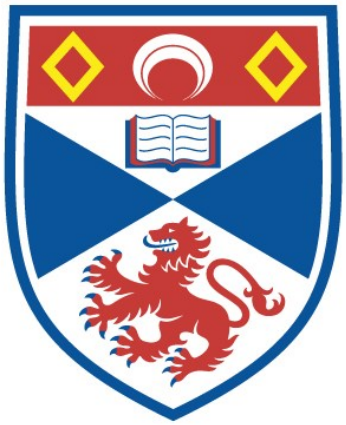
# LEGAL PATERNALISM

**Aislinn Batstone**

**A Thesis Submitted for the Degree of MPhil**
**at the**
**University of St Andrews**



**2003**

# Legal Paternalism

**Aislinn Batstone**

**Submitted in fulfilment of the requirements**
**for the degree of Master of Philosophy**
**at the University of St Andrews**
**18th August 2003**

## Contents

## Statement of Candidate

I, Aislinn Batstone, present this thesis for examination for the degree of Master of Philosophy at the University of St Andrews. I declare that the thesis is my own composition, that the work of which it is a record has been done by me, and that it has not been accepted in any previous application for any degree.

This thesis is a record of the work done as a research student in the period from September 2002 to August 2003, following my successful completion of the taught element of the degree in August 2002.

Access to this thesis in the University Library shall be unrestricted.

University of St Andrews

Departments of Logic & Metaphysics and Moral Philosophy
St Andrews, Fife KY16 9AL, Scotland

15th. August 2003

I hereby certify that this dissertation by Aislinn Batstone to the best of my knowledge fulfils Senate Regulations

Dr David Archard

Reader in Moral Philosophy

Mrs Janet Kirk, School Secretary
Tel: (01334) 462486/7
Fax: (01334) 462485
e-mail: philosophy@st-andrews.ac.uk

## Acknowledgements

My sincere thanks go first to the Commonwealth Scholarship Commission in the UK and the British Council, for the Commonwealth Scholarship that funded my research degree at the University of St Andrews. Also to my supervisor, David Archard, who inspired my interest in entirely new areas of philosophy, and to Sarah Broadie who fits encouragement and support for postgraduate philosophy students into her busy schedule.

I would not have got very far without my fellow postgraduates here at St Andrews. Thanks especially to the Barbers, and to the logically-minded Ross Cameron, Anne Manolakas and Marcus Rossberg for moral support.

Finally, I am very grateful to my family and friends outside Scotland for their support during the writing of this thesis.

Special thanks as always to my parents, for their love and patience.

## Abstract

Legal paternalism is the interference by the state in the choice or choices of citizens, for the good of those citizens. There are strong intuitions for and against paternalism, which correspond to the value we place on well-being and self-determination, respectively. This thesis takes as a starting point a certain balance between these two values, and explores its ramifications at a number of levels. Paternalism is taken to be justified only when the choices in question are not expressions of self-determination. I first explore what it takes for a choice not to be an expression of self-determination through a discussion of voluntariness and its relationship to compatibilism in discussions of freedom of the will. I conclude that which choices are expressions of self-determination is a matter to be settled in the practical domain. This idea is explored in a discussion of decision-making competence. Finally, I outline the positive normative ethical evaluation of self-determination which provides the justification for the anti-paternalist stance.

## Introduction

I have taken the problem of paternalistic law as the starting point for an exploration of ideas of freedom and responsibility at a number of levels. The structure of the thesis arises from the following crucial idea: that the question of when it might be justifiable for the law to intervene in individual's choices, for their own good, turns out to be unanswerable without consideration of issues in more than just the normative ethical field of philosophy. The outcome of the thesis is a clarification of interdependencies between metaphysical, ethical and political concepts, with the theme of paternalism providing both structure and an ever-present connection to pragmatic concerns.

In my first chapter I introduce the problem of legal paternalism and set out some common intuitions that have traditionally structured debate about paternalism. Intuition is the coin of moral philosophy, but its value in analytic circles depends on it being legitimised by concepts that have more theoretical depth. Joel Feinberg takes the first steps towards such a grounding of his anti-paternalistic intuitions, attempting to base them on a metaphysical distinction between voluntary and non-voluntary choices, and a positive ethical evaluation of individual sovereignty. However, I argue that the distinction between voluntary and non-voluntary choice is not as clear as Feinberg would like it to be.

This problem is taken up in my second chapter. Given the vast literature on metaphysics in this area, my discussion will of necessity be rather limited. But it will serve as a reminder of the importance of these issues to normative discourse, and a pointer for further research.
Various responses to the problem introduced in my first chapter are considered, but the most compelling one reverses the dependency of the ethical and the metaphysical: D.C. Dennett argues that our ethical practices should determine which actions we consider to be freely undertaken, rather than the other way around. On the basis of this idea, then, the choices in which we in fact intervene, will be those which are non-voluntary.

Putting the point bluntly like this exposes one great difficulty with it: that it seems to return us to the arbitrariness of competing intuitions, when it comes to the justifiability of paternalism (and blame, and other practical applications of attributions of responsibility.) My third chapter exposes this difficulty in detail through an exploration of the key manner in which paternalism is practically instantiated: determinations of decision-making incompetence. I argue that one *prima facie* plausible account of when we may 'decide for others' is in fact arbitrary in just this way, and thereby allows instances of paternalism which would not be tolerated by (for instance) Feinberg's more liberal theory of paternalism. I put forward one suggestion for reducing the role of intuition in competence determinations, but the fact remains that our intuitions even about why we may not interfere in a competent person's decisions, require a more solid grounding.

In my final chapter, therefore, I consider the *respect* for freedom that is the second element of any presumption against paternalism, seeking the normative basis for this respect. I consider particularly the difficulty of generating respect for the freedom of individuals at the political level, from normative ethical considerations: the problem of reconciling J.S. Mill's *On Liberty* and his *Utilitarianism* provides a classic illustration of this problem. In the end, though, I argue that Mill's basic ideas if not his precise formulation give us the right grounding for strong liberal intuitions against legal paternalism. Legal paternalism can be rejected on a consequentialist account that takes a certain sort of freedom as its goal.

## Chapter 1

### 1.0 Introduction

In discussions of the scope of the criminal law, legal paternalism is often thought to be one potentially viable exception to Mill's 'Harm Principle'. According to the Harm Principle, the *only* justification for legal intervention in the private lives of citizens is the prevention of harm to others. Mill's standpoint has become the touchstone for liberal (though not libertarian) accounts of the limits of the criminal law. Broadly, legal paternalism is the view that legal sanctions may also be justified in order to prevent people from harming themselves or consenting to being harmed by someone else. My project is to develop a position according to which legal paternalism is unjustified. This thesis is deep rather than broad, in the sense that I wish to develop a viewpoint on my position at a number of levels of philosophical debate, from the political to the metaphysical.

In this chapter I characterise paternalistic behaviour and paternalistic law, and set forth the guiding intuitions of the anti-paternalist stance. If proliferation is a mark of success, the theoretical landscape of this area is one of the success stories of reflective equilibrium: the same two basic liberal intuitions about paternalistic behaviour have made their way into a number of theories of what counts as justified, and what unjustified, paternalism. These intuitions are that a) in general, paternalistic intervention in the life of another person is not permissible, and b) paternalistic intervention *is* permissible where the person in question is a child, or not in their right mind, or dangerously ignorant. Different theories give slightly different analyses of when it is legitimate to interfere in someone's choices.

Joel Feinberg's is one careful and comprehensive example. Feinberg argues that voluntary choices are the expressions of an individual's sovereignty, which is never rightfully ignored. Thus in general, paternalistic intervention in the life of another person is not permissible. But non-voluntary choices, such as can be made in ignorance or mental incapacity, are not expressions of sovereignty. Therefore, paternalistic intervention can sometimes be legitimate. But Feinberg's theory faces one immediate metaphysical problem, namely in the difficulty of distinguishing voluntary choices from any other sort. We find the resolution of this problem

in the adoption of the practical stance advocated by D.C. Dennett. In a sense, freedom will be taken to be constituted by our moral practices.

## 1.1 Paternalism

Attempts to understand exactly what it is for an action or policy to be paternalistic suffer both from the negative connotations of the term, and from previous attempts at clarification which have muddied the waters. I will not give a survey of various characterisations of paternalism. I want, rather, to put forward a characterisation and extend it. In the process some other attempts at definition will be rejected. This characterisation aims to be relatively neutral in respect of the moral status of paternalism; in particular, paternalism is not taken as a *prima facie* wrong with no redeeming features. Nor am I concerned with paternalism from a particularly feminist perspective. The historical truth may be otherwise (the etymology of paternalism certainly misleadingly indicates otherwise), but *in principle* paternalism can be practised on *any*one, by *any*one else.

David Archard defines paternalism as follows:

> "P behaves paternalistically towards Q iff:
>
> (1)P aims to bring it about that with respect to some state(s) of affairs which concerns Q's good Q's choice or opportunity to choose is denied or diminished;
>
> (2)P's belief that this behaviour promotes Q's good is the main reason for P's behaviour;
>
> (3)P discounts Q's belief that P's behaviour does not promote Q's good." (Archard 1990 36)

Condition (1) – the denial or diminishment of choice – is a necessary element of a paternalistic action. A person may act for the purpose of promoting another's good, and do so against the opinion of that person that the act is *not* in their interest, and yet not behave paternalistically. For example, imagine a teenager who rebels against the wealth of his parents. He firmly believes that he is better off living by his own labour than accepting money from them. But his mother believes otherwise. She writes him a cheque for £ 500 000, and puts it in an envelope. She puts the envelope in a locker at the train station and gives the key to her son, telling him simply to open the locker if he ever changes his mind about needing money.

This (generous) mother has not interfered in her son's life beyond handing him a key, though she has acted in the belief that she is promoting her son's good, against his belief that she isn't. She has fulfilled conditions (2) and (3). But intuitively, her action is not paternalistic in the way that it would have been had she, say, packed him off to an expensive boarding school, thereby denying his choices. This brings up a further important point about paternalism on this characterisation: Parental acts will often be paternalistic. The younger the child, the higher the proportion of paternalistic acts (the fewer choices the child is allowed to make for herself.) This is a further reason to note that paternalism can not be universally wrong, although this first condition regarding the denial or diminishment of choice will be part of the case against it.

A paternalistic act must clearly also aim to bring about the good of the person concerned. One who aims simply to bring it about that another's choice is diminished, discounting that other's opinion that the interference is not in their interest, is merely a malicious intervener. The parental aspect of the etymological root of 'paternalism' is here relevant: that it be for its object's good is a necessary element of paternalism, and provides it with positive flavour for its advocates. To illustrate, picture a woman who one evening locks her protesting husband in the bathroom and leaves the house for a night out with friends. She denies her husband's choice to have the run of the house while she is out, against his opinion that incarceration is *not* in his interest. Her action lacks the benevolent undertone necessary to an act of paternalism. If she had locked him into the bathroom so that he couldn't get to the gun with which he'd threatened to shoot himself (but still against his view of his own best interest, which he considers is to die), and left to get help, then her action could be considered paternalistic.

Finally, consider the last clause. The paternalist must act against what the person affected takes to be her own best interest. This clause counteracts the possibility that the person affected has *chosen* her diminishment of choice. The classic case of Odysseus and the sirens presents some difficulties. Odysseus instructs his crew to bind him so that he will not be tempted by the voices of the sirens. Though he might beg them to release him, they must not – his earlier choice is to overrule the later. In that their actions are in accordance with his overall choice, the crew are not behaving paternalistically. But in that they are denying his *current* choice (against what he *currently* takes to be his best interest), they *do* appear to be behaving paternalistically. What we need to show the necessity of the third clause is not

something of the 'don't let me...' form. In these cases *choice* cannot be separated from a person's opinion as to their best interests in the manner required: choice at a moment will always track opinion as to best interest at a moment. It is therefore ambiguous as to whether the intervention is paternalistic in these cases.

A better example case is one where a person becomes definitely committed to some desirable course of action only if some antecedent condition is met. If Angela were offered a job in Rome, for instance, she would be bound to take it, because she loves Italy and has always wanted to live there. It would also be wonderful for her career as a curator. If Amy pulls strings to get Angela a job offer from a museum in Rome, she has diminished Angela's choice – Angela is now bound to go and live and work in Rome. We can also presume that the act was for Angela's benefit, and not because Amy wanted to get rid of her. But this act lacks the *mal*evolent undertone of paternalism – intuitively, since it is not against what *Angela* takes to be her own best interest, it is not paternalistic. If Amy had had Angela appointed at a museum in Rome for the good of her (Angela's) career, but *against* her desire to stay in Scotland with her family, then Amy's string-pulling would be paternalistic.

An act that meets all three conditions *is* paternalistic. The mother who packs her protesting son off to boarding school so that he will be properly educated behaves paternalistically, as does the wife who locks her suicidal husband in the bathroom. If Amy sends Angela *protesting* to a better life in Rome she behaves paternalistically. It should be clear from these examples that the ethical status of paternalism is not straightforward.

Archard's definition differs from its predecessors in a number of respects. Firstly, it does not focus on the restriction of the *freedom* of action of the agent who is the object of paternalistic behaviour. A conception of paternalism as the restriction of freedom of action faces the following counter-example. Suppose a doctor decides to give an unconscious patient a life-saving blood transfusion, when she knows he would reject the transfusion if conscious. The doctor's behaviour seems to be paternalistic but not restrictive of her patient's freedom. He is unconscious, and so not free to act anyway. (Archard 1990 36; Gert & Culver (G&C) 46) Thus restriction of freedom of action cannot be the defining feature of paternalistic behaviour.

Nor, Archard argues, does paternalistic behaviour necessarily involve doing something which "needs moral justification", as conjectured by Gert and Culver. They suggest that "an essential feature of paternalistic behaviour toward a person is the violation of moral rules … for example, the moral rules prohibiting deception, deprivation of freedom or opportunity, or disabling." (G&C 48) But the breaking of a variety of rules doesn't seem to be a necessary feature of paternalistic behaviour. Archard cites an example of Gerald Dworkin's: "A husband is under no moral obligation to inform his suicidal wife of the whereabouts of *his* sleeping pills. He acts paternalistically when he fails to do so, but violates no moral rule by his failure. [...] The husband seeks to diminish the wife's opportunity of choosing suicide by denying her access to the pills." (Archard 1990 37)

This is actually a somewhat tricky case, since the husband is failing to act rather than actively intervening. If his wife does not know that he has sleeping pills, and he neglects to tell her, he does not seem to be behaving paternalistically on Archard's model either. The wife's range of choices never included overdosing on her husband's sleeping pills, and he has not diminished this range of choice. Suppose, then, that she knew of the existence of the pills, asked him where they were, and he refused to tell. Then his behaviour seems to be paternalistic. But it seems quite right to balk at saying that he has violated a moral rule, though some question remains as to whether his action was wholly, morally unproblematic. If his wife is capable of making a rational decision then there is at least some form of disrespect involved in his action.

Archard distinguishes paternalistic behaviour from paternalistic *reasons* for acting. A paternalistic reason is "a belief that a behaviour promotes the good of some other, together with a belief that this first belief is not shared by the other, and of course a desire to promote the good of the other." (Archard 1990 38) The idea of a paternalistic reason for behaviour allows us to capture the paternalistic element of behaviours that fail to fully meet the terms of the definition. Consider again the mother who sends her errant son to an expensive boarding school in another town. But suppose this time that the son doesn't *really* want to rebel against his parents' wealth: he was deviously acting rebelliously in order to get sent away. He wants to be wealthy, but also to have the freedom of living away from his parents. Then we might be hesitant about saying that the mother behaves paternalistically. She hasn't acted against her son's belief regarding his own good, though she thinks she has. But we can still say that she acts for paternalistic reasons. This distinction is made largely to keep the definition of

paternalism clear and precise, since it is debatable whether the mother behaves paternalistically according to common usage of the term.

## 1.2 Legal paternalism

Archard's conception of paternalism seems to conform to prephilosophical ideas with respect to its classifications of which behaviours are paternalistic and which aren't, and is precise enough to provide solid ground for the discussion ahead. We may need to make some modifications to extend this tidy definition of paternalism to the law. Condition (1) can be left relatively intact. Where Q stands for any citizen of a particular state:

> (1a) The state aims to bring it about that with respect to some state(s) of affairs which concerns Q's own good Q's choice or opportunity to choose is denied or diminished.

We can, perhaps, legitimately talk about the 'beliefs' of the state, as shorthand for the consensus of expert opinion available to policy-makers. So, for instance, the state may 'believe' that smoking causes lung cancer. The state's 'belief' that smoking causes lung cancer prompts certain policies and programs: it may be part of the basis for high taxes on cigarettes, and for advertising campaigns warning of the dangers of smoking. We can thus modify Archard's condition (2) for the state:

> (2a) The state's belief that this policy or law promotes the good of Q is the main reason for the implementation of that policy or law.

The tricky part of this definition of paternalism for the law is condition (3), because the force of the law extends to an entire population. Sometimes the individuals affected will be of the opinion that the law is in their interests, sometimes they won't. Most people, for instance, feel that wearing a seatbelt is in their own best interest. But there may be a minority who disagree, either because they value their freedom over their health, or because they don't believe seatbelts actually improve their safety. We can thus reformulate condition (3) as follows:

> (3a) The state discounts any belief Q might have that its law or policy does not promote Q's good.

This allows us to count as paternalistic laws which are not strictly paternalistic in every application, such as the law on seatbelt use.

## 1.3 The intuitive cases for and against

*Prima facie*, there are many cases of paternalistic interference at the personal level that appear ethically questionable to anyone with a robust notion of personal autonomy. For instance, the medical profession has learned in recent years that patients resent having decisions about their treatment taken for them, regardless of the years their doctors have spent studying medicine – in other words, regardless of the fact that doctor really does often know best. The doctrine of informed consent is the result of this basic reaction against paternalism. The reason the word has such negative connotations is that people value making their own choices, and have firm beliefs about what is and is not in their best interests. Paternalism, though it is motivated by the desire to help, ignores those choices and the beliefs regarding them.

However, sometimes paternalism (as I have argued it should be characterised) does seem to be justified. The simplest case is probably that of young children. For an act to be paternalistic at all, the subject must be (or have been) capable of choice and belief – otherwise the conditions of denial or diminishment of choice and ignoring the beliefs of the subject cannot be met. So the acts parents make on behalf of babies and very young children cannot be regarded as paternalistic. But once a child is old enough to make choices and have beliefs about what is in his interest, he is old enough to have those choices denied and those opinions disregarded.

This sort of paternalistic behaviour is not usually as malicious as it sounds on this abstract characterisation. Every time a parent gives a child a piece of fruit instead of the sweet she wants and thinks best for her, he is behaving paternalistically – but it seems to be warranted. Something about the quality of her choices and beliefs (which we will come to later) seems to warrant denying the one and ignoring the other. Similarly, a psychiatric patient may legitimately be detained at a hospital if she is judged to be an immediate danger to herself. Such an allowance seems reasonable, as does the behaviour of a person in preventing his drunken friend from stepping out into traffic. Paternalism seems outrageous when it is aimed at capable adult decision-makers; in many other cases, intuitively, it may be legitimate.

But in the legal domain this modest view has not always prevailed. The 'one-part' case is one in which an individual simply wishes to act in such a way as to harm herself – for instance,

by using dangerous drugs. The 'two-party' case is one where an individual requests or contracts another to commit the injury upon her. Joel Feinberg formulates the distinction between what he calls 'soft' and 'hard' legal paternalism as follows:

> "The distinction, which is of the first importance, has to do with the weight attached to the voluntariness of a person's action in the one-party case and the voluntariness of his consent in the two-party case. Hard paternalism will accept as a reason for criminal legislation that it is necessary to protect competent adults, against their will, from the harmful consequences even of their fully voluntary choices and undertakings. [...] It is not as clear that "soft paternalism" is "paternalistic" at all, in any clear sense. Certainly its motivating spirit seems closer to the liberalism of Mill than to the protectiveness of hard paternalism. Soft paternalism holds that the state has the right to prevent self-regarding harmful conduct (so far it *looks* paternalistic) *when but only when* that conduct is substantially nonvoluntary, or when temporary intervention is necessary to establish whether it is voluntary or not" (Feinberg 12).

On the characterisation of legal paternalism given above, what Feinberg terms 'soft' paternalism *is* paternalism: the state aims, by law, to diminish the choices of individuals, for their own good, regardless of their view of their own interests. It is just that the *choice* involved, in the case of 'soft' paternalism as opposed to 'hard', is taken to be less than fully voluntary.

So how is Feinberg's soft paternalism "closer to the liberalism of Mill"? Mill's Harm Principle is the classic statement of the scope of the law; it says that the state is justified in interfering in the private lives of citizens *only* to prevent harm to others. The law may *not* prevent a competent adult from harming himself: "His own good, either physical or moral, is not a sufficient warrant" (Mill *On Liberty* 15). Mill's view contrasts what Feinberg would call 'hard' paternalism, the view that the state may legitimately prevent competent adults from harming themselves.

Mill goes on to write that complete freedom from intervention is only applicable to "human beings in the maturity of their faculties" (15). We *are* justified in protecting children, for instance, from their own harmful choices. Further, we are allowed to intervene in order to ascertain whether a person knows what he is doing, if the harm is imminent. Mill's famous example of the dangerous bridge sets out his views on the justifiability of various forms of intervention:

> "it is a proper office of public authority to guard against accidents. If either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall in the river. Nevertheless, when there is not a certainty, but only a danger of mischief, no one but the person himself can judge of the sufficiency of the motive which may prompt him to incur the risk: in this case, therefore (*unless he is a child, or delirious, or in some state of excitement or absorption incompatible with the use of the reflecting faculty*), he ought, I conceive, to be only warned of the danger; not forcibly prevented from exposing himself to it" (Mill 118 [my emphasis]).

If Feinberg's account of the 'non-voluntariness' of certain choices agrees with Mill's (italicised) exceptions to the general rule of non-interference, then Feinberg's soft paternalism is indeed in the spirit of Mill's liberalism. *Prima facie* the two do agree: choices that Feinberg classes as non-voluntary are those made when the "use of the reflecting faculty" is somehow severely diminished – by ignorance, illness, drugs, or insufficient maturity, as well as those that are more clearly forced. They are not the ""Unreasonable choices"...commonly made by fully competent persons in full command of their faculties" (Feinberg 106). We will look more carefully at what conditions contribute to the non-voluntariness of choices later on.

Most of the laws that prevent people from inflicting harm upon themselves seem to arise from a number of motives, or at least can be justified from a perspective other than paternalism, if necessary. For example, the view that the wearing of seatbelts should be legally enforced may be a hard paternalist stance – the law does diminish the choice of individuals against (some of) their judgement of their own best interests. The stance is hard paternalism if the motive is in fact the protection of the interests of those individuals. But if the motive is to protect the bulk of taxpayers from the jointly-felt medical burden of foolhardy drivers, the law is not paternalistic. Probably, the law is a mixture, since a state (as a conglomerate of individuals with varying interests and levels of power) can rarely be said to have a single motive for any of its policies.

Similar complexities arise in debates about the legalisation of cannabis and harder drugs. On the one hand, if the law is in place to protect citizens from their choices regarding these

drugs, it is paternalistic. But it may be grounded in a desire to protect the public from the misbehaviour of drug users, or to protect children from potential drug use problems, or it may be based on moral disapprobation of drug use. In these cases, rather than hard paternalism, the law is based on the Harm Principle, soft paternalism, or legal moralism respectively. It is only where a law aims to protect capable decision-making adults from their own choices that it falls into the category of hard paternalism, but such laws are by no means unheard of.
It is difficult to think of examples of laws based on the principle of *soft* paternalism, perhaps because, as Feinberg writes, the principle is more naturally seen as a *restriction* on legislation. It says that "a certain class of alleged justifying reasons are *not* valid. It is *not* an acceptable reason in support of proposed criminal legislation that it is necessary to prevent the sorts of interferences soft paternalism permits. Interfering with an apparently demented suicide attempt, for example, should not be a crime" (Feinberg 15). Soft paternalism also legitimates the detention of some psychiatric patients, as mentioned above. Also, the principle of soft paternalism is more likely to be the basis for non-punitive state interference with liberty: "denying applications, invalidating contracts, issuing temporary restraining orders, imposing civil commitment, and so on" (15). But as mentioned earlier, soft paternalism may be part of a mixed justification for laws governing sexual conduct, drug use and freedom of speech. Some restrictions in these areas might be argued for on the grounds of protecting children from the harmful consequences of their own choices (though whether this is a good argument is another matter).

### 1.4 The presumptive case against hard paternalism and for soft paternalism

Mill and Feinberg both argue against a hard paternalist approach in the law, opting instead for the soft paternalist stance. The presumptive case against hard paternalism is not based solely on respect for a person's choices, but on respect for autonomy, self-determination, or sovereignty. Mill's argument against hard paternalism is based on a number of factors, one of them being the wider value of individuality – more on this later. But Feinberg advocates that we use an analogy between persons and sovereign states to get to the heart of what is wrong with hard paternalism. Sovereignty (in the political sense) consists in a right of self-determination (Feinberg 47). According to Feinberg, sovereign states may have important and rewarding relationships with other states (in contrast with the conception of autonomy as mere *independence* that, for example, O'Neill discusses) but they are essentially self-governing within their own domain.

Feinberg writes:

> "Perhaps the fairest way of putting the presumptive case against legal paternalism is to say that even when conjoined with other principles, it has at best a very limited conception of personal autonomy. Even though it is consistent with the recognition of a person's right of self-determination, it subordinates that *right* to the person's own *good*" (Feinberg 57).

What is a person's good?

> "A majority view, associated with the writings of Plato, Aristotle, Rousseau, Hegel, and Mill, among others, identifies a person's good ultimately with his *self-fulfilment* – a notion that is certainly not identical with that of autonomy or the right of self-determination. self-fulfilment is variously interpreted, but it is usually understood to require the development of one's chief aptitudes into genuine talents in a life that gives them scope, the unfolding of all basic tendencies and inclinations, both those that are common to the species and those that are peculiar to the individual, and the active realization of the universal human propensities to plan, design, and make order. self-fulfilment, so construed, is not the same as achievement and not to be confused with pleasure or contentment, though achievement is often highly fulfilling, and fulfillment is usually highly gratifying" (Feinberg 57).

Self-fulfilment and the right to self-determination are clearly closely linked. Health and material and psychological welfare are also related to self-fulfilment, probably instrumentally so.

Feinberg lists four possibilities as to exactly how self-determination and self-fulfilment are linked. The first view is that self-determination is valuable *solely* in terms of its being instrumental to self-fulfilment. How does self-determination work towards a person's achieving self-fulfilment? There might be more than one mechanism. But primarily, a person may be taken to have greater knowledge of what will fulfill them than anyone else will. Being self-determining will allow them to apply this knowledge to the maximisation of self-fulfilment. This is the view most naturally linked to paternalism: as soon as self-determination *stops* being instrumental to a person's good, it may be disregarded. So, for instance, if a person's self-determination clashes with their health, self-determination may be sacrificed in order to preserve health, and therefore the overall good of the person, or their ability to achieve self-fulfilment in the future. But an anti-paternalist with this view of the

relationship between self-determination and self-fulfilment might stress the instrumental *importance* of self-determination for self-fulfilment (Feinberg 59). No-one can know a person's good better than they can, and even being in perfect health won't help them if they cannot achieve what they want in other areas.

If the stress put on the instrumental value of self-determination is heavy enough, we shift to the second view, where "the relation between a person's right of self-determination and his good of self-fulfilment is not merely a strong instrumental connection but an invariable correspondence. On this view, whatever harm a person might do to "his own good" by foolishly exercising his free choice would in every case necessarily be outweighed by the greater harm done by outside interference and direction" (Feinberg 59). Self-determination is still instrumental to self-fulfilment, but it has become an overriding consideration: its instrumental value is such that it overcomes other considerations. However, if the value of self-determination is truly only instrumental to self-fulfilment, it is hard to see how its value could overcome *all* other considerations. If everything required to make a person fulfilled could be brought about without her effort, she would be just as fulfilled as if she had determined the outcome herself. This is what it means to say that self-determination is instrumental to self-fulfilment. Thus, on the instrumental views, hard paternalism would be justified if it created an increase in the overall good of the person large enough to offset the loss of self-determination.

The third and fourth views, according to Feinberg, take the right of self-determination to be entirely basic. Self-determination is not instrumental to self-fulfilment or even a part of it. It is something to be valued independently of self-fulfilment. Here, we have the potential for clashes between self-determination and self-fulfilment. On the third view, self-determination will always win in these clashes: it will over-ride self-fulfilment. Here there is clearly no room for hard paternalist intervention: the denial of self-determination can never be legitimate. On the fourth view self-determination and self-fulfilment are to be *balanced* against one another, allowing some scope for paternalistic considerations when some important aspect of a person's good may be damaged by their choices (Feinberg 60).

It is the third view that Feinberg finds consistent with his model of personal sovereignty, and that he prefers.

> "When the exercise of a person's sovereign right conflicts with what is truly good for him (those "rare cases"), [my view] defends the choice nevertheless. If that seems an absurd result, the reader should put himself in the position of the person interfered with. Presumably, if he genuinely chose the alternative that is in fact bad for him, he did not choose it because *he* believed it was bad for him. That would be so irrational that it would put the voluntariness of his choice in doubt, and...the soft paternalistic strategy might then be used to justify interference on liberal grounds. If he chose that alternative because *he* believed it good (or at least not bad) for himself, then either the difference between him and his would-be constrainers is over some matter of fact about which he is simply mistaken, in which case he would welcome being set right, or it is about the nature of his self-interest, or the reasonableness, given his values, of the risks he wishes to assume. In that case, the disagreement would be more intractable, and the reader would not welcome having his own judgment overruled, or the "better values" of others substituted for his own" (Feinberg 62).

We can also conceive of a person sacrificing her own good for the sake of others, or some "treasured cause", or deliberately valuing a short-term gain over a long-term good. To the extent that these choices are truly voluntary (as we shall see), they are to be respected as expressions of the right of self-determination.

Once we have the grounding for the presumptive case against hard paternalism in place, it is easy to capture the justification for Feinberg's 'soft' paternalism. Paternalism, we have seen, consists in the denial or diminishment of choice. What is wrong with it is that it violates the right of self-determination. But if a certain choice is not an expression of the right of self-determination then the paternalism is not presumptively wrong. What choices are not expressions of the right to self-determination? According to Feinberg, substantially *non-voluntary* choices can be denied without violating the right to self-determination. 'Soft' paternalism will allow interference in these choices.

## 1.5 Voluntary and non-voluntary choices

The distinction between fully voluntary and non-voluntary choices is crucial to Feinberg's account of what is *justified* legal paternalism. It is the key to his 'soft paternalist strategy': Feinberg uses the idea of substantially non-voluntary choices to ground his view that the

paternalistic interferences that look reasonable are in fact justifiable on liberal grounds. Interference in a choice that is substantially *non-voluntary* is significantly different from interference in a choice that is *voluntary*. The difference is significant enough that the denial or diminishment of non-voluntary choices may be justified, while denying voluntary choices is never legitimate.

The natural beginning of a discussion about voluntariness is with Aristotle. Aristotle's conception of voluntariness is notoriously strict – that is, he classes as voluntary any action which "has its first principle in the person himself when he knows the particular circumstances of the action" (NE III 1111b). The two factors that contribute to an action's failing to be voluntary are thus *ignorance* and *force*. Aristotle has a rather narrow conception of force: a person is forced to act only if their body is literally externally moved (NE III). "What is forced, then, seems to be what has an external first principle, where the person forced contributes nothing" (NE 1110b). Being moved by a force of nature such as a flood or hurricane is an example of non-voluntary action on Aristotle's account, or being physically moved by another person. But that person threatening your life if you fail to move does *not* count as a mitigating factor against voluntariness. "In fact, the person acts voluntarily, because in actions like this the first principle of moving the limbs that serve as instruments lies in him; and where the first principle lies in a person, it is in his power to act or not to act" (1110a).

Though the mitigating factors of ignorance and force seem still to be a useful way of understanding what it is for an action to be non-voluntary, most modern writers would expand Aristotle's conception of force. In particular, we might want to allow that there may be *internal* sources of force, and external sources of force which are not purely physical. Thus Feinberg distinguishes internal from external sources of compulsion; internal sources include "Neurotic compulsions, obsessions, inhibitions and incapacities" (Feinberg 151). These are all factors that will tell against the voluntariness of an individual's action. Under the heading of external non-physical compelling forces, we may wish to include brainwashing and hypnosis along with coercion. So non-voluntary choices will be those made under the influence of drugs, or agitation, or certain mental illnesses, or when a person is insufficiently mature – in other words, in complex cases of ignorance and internal force.

Note that as soon as these internal elements of force are added (if not before) the boundary between voluntary and non-voluntary actions starts to become blurred. Some actions performed (or choices made) under conditions of internal or external non-physical force we may want to say are completely involuntary, akin to being swept down a river. But, for example, in the case of coercion we can clearly see a continuum between non-voluntary and substantially voluntary (i.e. voluntary enough for the purposes of the law) choices. Coercion, Feinberg writes, closes off alternative possibilities to a person by increasing their cost to an unacceptable level ("Your money or your life!") (Feinberg 192). Typically, threats are made in order to drive the coercee to a particular action. Alternatively, an *offer* may be coercive ("I will pay off your debts if you sleep with me"). Handing over your money to the gunman is a clear case of a non-voluntary choice – though there was an alternative (your life), it was unacceptably costly. But if you are told "Sleep with me or I will be annoyed!", the decision to acquiesce (barring other factors) would appear to be substantially voluntary. In the case where the choice turns out to be substantially non-voluntary on Feinberg's account, it is not thought literally to be the result of psychological compulsion – it is rather that the alternative has been made unacceptably costly.

Why might it be all right to interfere in a person's substantially non-voluntary choice? It might be argued that substantially non-voluntary choices are really like *someone else's* choices, and so the person who is harmed by their own substantially non-voluntary choice is really harmed by someone else. The drunken friend is a different person to the friend while sober. Pulling the drunken friend out of traffic is just stopping him from harming his alter ego, the sober friend. Thus soft legal paternalism would really fall under the Harm Principle. Such cases as the legal allowance of detention in a psychiatric ward would count as preventing the suicidal patient from harming another (her later self), just as the detention of a dangerously psychotic patient is allowed on the grounds that he may cause harm to other people. But Feinberg rightly denies the "absurd fiction" that "such factors as ignorance are themselves "other persons" who can be targets of legal threats" (Feinberg 14). If the foregoing model is taken seriously, the suicidal patient may not only be restrained, but also be punished for any harm she might inflict on 'herself' – since this is really harm to another. But this is absurd. There is no second person who may be sanctioned for her behaviour against the first. It cannot, therefore, be the case that soft paternalism simply falls under the Harm Principle.

Feinberg's account of substantially non-voluntary choices – those that may rightfully be interfered with – is based on a comparison with the antithesis of perfect voluntary choice: involuntary action. An involuntary act is one that literally can't be helped; it is externally forced on the Aristotelian model. In fact, involuntariness and choice are conceptually incompatible. Unlike Aristotle, Feinberg holds a relatively small class of actions to be fully voluntary. They are those that are made under ideal conditions: the agent has full knowledge of current circumstances and possible outcomes, and suffers from no failure of cognition or will. Needless to say, Feinberg's ideal of perfect voluntariness is practically unattainable. But choices can lie on a scale between perfect voluntariness and involuntariness. Substantially non-voluntary choices appear to *approach* (in the mathematical sense) involuntariness (Feinberg 104-105).

For Feinberg, what is wrong with paternalism, when it is wrong, is that it is a violation of sovereignty – a person's right to self-determination. But interfering with someone's involuntary action clearly does not infringe on her right to self-determination. The act is not even chosen; it cannot, therefore, be an expression of self-determination. Insofar as non-voluntary choices approach involuntary choices, then, neither are they expressions of self-determination. It is necessary for Feinberg that the 'non-voluntary' choices made while drunk, 'neurotic', and so on, are seen as approaching involuntary 'choice': this provides a strong justification for his soft paternalism. It is justified by the fact that it isn't interfering with true self-determination at all. For Feinberg, personal sovereignty is a trumping consideration – we aren't justified in violating sovereignty, even when a person's life is at stake. To justify any sort of intervention, then, Feinberg must argue that the choice at hand does not fall in the protected domain of sovereignty.

But note that Feinberg is wrong to attempt to *contrast* soft and hard paternalism: soft paternalism, though it may be valid where hard paternalism is not, is nonetheless paternalism. Soft paternalism's characterisation as such (and not as "anti-paternalism") can be based on its fulfilling the criteria for paternalism set out above. When some policy is made with the intention of protecting people from their own substantially non-voluntary choices – for instance, if poker machines are banned from pubs (presuming that patrons will be under the influence of alcohol when they use the machines) – all the criteria for paternalism are met. It seems that the state aims to bring it about that with respect to some state of affairs that concerns Q's own good Q's opportunity to choose is denied (Q can no longer choose to

gamble while she drinks). The state's belief that this policy promotes the good of Q is the main reason for the implementation of that policy, and the state discounts any belief Q might have that the policy does not promote Q's good. The first condition is clearly the crucial one: the crux of paternalism is the usurpation of choices. Where an act is involuntary, there really is no choice made at all. In cases where non-voluntariness is practically indistinguishable from involuntariness, we may wish to say the same of non-voluntary 'choices'. But in other cases of non-voluntary choice – even *substantially* non-voluntary choice – it must be admitted that a *choice* really is being made. It may not be the choice that *would* be made in a calm, reflective frame of mind, but as a choice its denial or diminishment is paternalistic.
One last thing we should note about Feinberg's account of the voluntariness/non-voluntariness distinction is that standards of voluntariness – how voluntary a choice must be before it is immune from paternalistic interference – may vary. In particular, standards of voluntariness will vary with the riskiness of the activity to be undertaken. "Most harmful choices, like most choices generally, fall somewhere in between the extremes of full voluntariness and complete involuntariness. It follows that we may formulate relatively strict (high) standards of voluntariness or relatively low standards of voluntariness in deciding, in a given context and for a given purpose, whether a dangerous choice is voluntary enough to be immune from interference" (104). In practice, the outcome of this qualification is that the riskier the action to be undertaken, the greater the degree of voluntariness that will be required of the chooser. Further, the more permanent or irrevocable the potential damage, the greater the degree of voluntariness the person making the decision will need to display (119-120). If the person falls short of these higher standards of voluntariness, it will be legitimate to paternalistically interfere in their choice.

## 1.6 Conclusion

There is a conceptual distinction between involuntary acts and voluntary choices. Involuntary acts cannot be thought of as the outcome of choice at all, while voluntary acts, though they may fall short of the ideal of a perfect understanding of consequences and a calm mental state, are the result of choices we make. But, intuitively, the more the voluntariness of an act falls short of this ideal – the more factors we are ignorant of, the more agitated we are – the closer our choice comes to being involuntary.

The second part of Feinberg's story introduces the normative notions required for a distinction between justified and unjustified paternalism. The value introduced is personal sovereignty, which Feinberg argues is a trumping consideration. Our right to personal sovereignty wins over all other considerations in judgements about our good.

Involuntary acts are certainly not expressions of sovereignty, or self-determination. I do not even choose to breathe, or for my heart to beat. I do not choose to stumble into traffic, and the person who pulls me out of danger is not violating my sovereignty. There is nothing wrong with what he does. Insofar as choices approach involuntariness, then, there is nothing wrong with interfering in them. (In fact there may be everything right with interfering in them, as in the traffic case.) Thus Feinberg uses the conceptual distinction between involuntary and voluntary acts, and the continuum postulated between them, along with an argument for the value of sovereignty, to ground an ethical distinction between justifiable and unjustifiable paternalism. Interference in voluntary choices is unjustifiable, while the usurpation of choices that are substantially non-voluntary is justified. In the second chapter we look more closely at this distinction between voluntary and non-voluntary choices.

## Chapter 2

### 2.0 Introduction

In the previous chapter I set out some crucial distinctions, provided solid definitions to ground the discussion to follow, and began to explore some of the difficulties in determining the justifiability of paternalism. I extended Archard's basic definition of paternalism to the legal context, and went on to look at Feinberg's distinction between soft and hard paternalism. For Feinberg, hard paternalism involves the usurpation of even voluntary choices, where soft paternalism is the denial or diminishment of substantially non-voluntary choices. Hard paternalism is presumptively in need of justification where soft paternalism is not, because hard paternalism is disrespectful of the self-determination of the individual, where, on Feinberg's account, soft paternalism is not. Soft paternalism involves the usurpation of choices that are not properly thought of as expressions of self-determination.

In this chapter I outline some problems with Feinberg's characterisation of soft and hard paternalism in terms of non-voluntariness. To speak of coercive 'compulsion' or mental 'pressure' is to speak metaphorically, generating puzzles about identity and responsibility. How do we make a distinction between voluntary and non-voluntary choices, when, firstly, both can be made by the same person in the same state and, secondly, the sceptic says there are no voluntary choices at all? The distinction between voluntary and non-voluntary choices that Feinberg needs to make to ground his position needs a metaphysical basis. I reject one account of such a basis and tentatively accept another.

Feinberg's theory of paternalism and when it is justified has two major aspects, the distinction between voluntary and non-voluntary choices and the value that is placed on individual sovereignty. In this chapter we are questioning the basis of the first aspect, which takes us into issues of metaphysics. Though problems of moral and criminal responsibility are in recent times usually discussed without reference to such issues, I believe it is useful to be reminded of their importance in grounding any coherent philosophical ethic. However, given the complexity of the metaphysics of freedom (and the vastness of the philosophical literature in this area) this discussion does not aim for firm answers. It is intended rather to flag the relevant issues and to be a pointer for further research. The ensuing chapters will take us into

a discussion of the second aspect of Feinberg's theory, the value that is put on individual sovereignty or self-determination.

## 2.1 The Van Gogh Problem

Let's now take a closer look at the conceptual distinction between voluntary and non-voluntary choices. *Involuntary* acts are the model for Feinberg's 'non-voluntary' choices, and the basis for his claim that paternalistic interference in non-voluntary choices is justified. What is involuntary is not even a matter of choice; it is pointless to desire control over the involuntary, and arguing for the value of self-determination in the sphere of the involuntary would be absurd. Feinberg presents non-voluntary choices as analogous to the involuntary, though the fact that a *choice* can be non-voluntary indicates an immediate disanalogy. In likening non-voluntary choices to the involuntary Feinberg is making use of the fact that we think metaphorically of coerced and irrational acts (classically non-voluntary) as forced. It is this metaphorical link that sanctions the step from the justifiability of intervention in *in*voluntary acts to the justifiability of intervention in *non*-voluntary choices.

However, the fact that non-voluntary choices aren't literally compelled in the same way that involuntary choices are creates problems for Feinberg's view. We can imagine Van Gogh, at least at some stages, feeling the same sense of compulsion to paint as he did to mutilate his own ear. Both the artistic expression and the self-mutilation on Feinberg's account would seem to be substantially non-voluntary choices, if Van Gogh is in the same mental condition when both acts occur; Van Gogh's self-mutilation seems to be an uncontroversial case of non-voluntary action. But the idea that paternalistic denial of the choice to paint mad but beautiful pictures can be justified is ludicrous. On the other hand, there may have been justification for preventing the self-mutilation.

It is not simply that there is necessary but not sufficient cause for intervention in the case of painting. Certainly, we usually require there to be an element of danger to a person before we intervene paternalistically ("for his own good"). Given the importance of painting to the *identity* of the artist, paternalism seems simply unjustifiable. Van Gogh wrote to his brother "And really, as for the artist's madness…I do not say that I especially am not affected through and through, but I say and will maintain that our antidotes and consolations may,

with a little good will, be considered ample compensation" (Van Gogh, quoted in Cabanne 169).

The consolation is in the feeling of pride the artist can take in his work. The flip side of freedom is responsibility – not only moral responsibility, but the taking of responsibility for one's own life and choices that creates an individual character. At least in some important cases, such as the Van Gogh case, for an action or choice to count as character-constituting, the individual must be responsible for it. But a choice for which a person can truly be said to be responsible must be undertaken voluntarily. Thus for Van Gogh's painting to count as character-constituting, it must have been undertaken voluntarily. The problem with Feinberg's argument by analogy is that while the slide from involuntary to non-voluntary is useful in some cases – where responsibility really does seem to be diminished – in other cases it seems premature and potentially dangerous to the identity of those to whom we apply it. The problem is that there are cases – such as the Van Gogh case – where we want to attribute responsibility for one act that is committed in the same mental state as another act for which we do not want to attribute responsibility.

A severely depressed person is not literally compelled to suicide. She wants to end the pain, and she believes that death will end the pain. She is right. That doesn't mean this is the best choice she could make. There are some cases where it seems very clear that some level of paternalistic interference is warranted and justified – where, for instance, the depression is to a large extent treatable and the patient fails to comprehend this fact. On Feinberg's model this intervention is justified on the grounds that the depressed person's choice to commit suicide is *non-voluntary*. Feinberg makes use of a metaphorical link between his 'non-voluntary' and *in*voluntary acts. Interference in the latter is clearly not a violation of self-determination. By analogy, interference in the former is likewise justified. But the crucial difference is that choices that seem to be non-voluntary (because they are committed in exactly the same mental conditions as uncontroversially non-voluntary choices) are real and sometimes important choices of that person – they can have a bearing on identity in a way that involuntary acts cannot. The likeness between involuntary acts and non-voluntary choices fails to hold up under analysis, and the step from the justifiability of interference in involuntary acts to the justifiability of paternalistic interference in non-voluntary choices seems to inherit its difficulties.

## 2.2 A possible response

There is a fairly standard response to this problem. Feinberg himself outlines one form of it in an early paper, 'Causing Voluntary Actions'. If Van Gogh's illness causes him to mutilate his ear, but also causes him to paint beautiful pictures, the puzzle is how one action can be voluntary (as it seems it must be if we are to hold him responsible for it, or praise him), while the other is non-voluntary.

Feinberg, it seems, would agree that the action of painting beautiful pictures is voluntary. But the notion of voluntariness is not quite as strong as might have been supposed. Feinberg's article is concerned with external human causes – that is, cases where one person causes another to commit some voluntary act. One illuminating and simple example is the following: "I can get an acquaintance to say "good morning" by putting myself directly in his line of vision, smiling, and saying "good morning" to him. My doing these things is not only a circumstance but for which his voluntary action would not have occurred, it is also a circumstance which, when added to those already present, "made the difference" between his speaking and remaining silent" (Feinberg 1970 177). The acquaintance's greeting seems to be voluntary, by any account, and yet, according to Feinberg, my action is a sufficient cause of it. However, on Feinberg's account there are still actions that are non-voluntary *because* they are caused, either by some abnormality within the agent, or by someone or thing outside the agent. Failures of voluntariness can arise through insanity, coercion, ignorance and so on – all these factors can be said to cause the act (1986, 115; chapters 21-26), and when they do so, the act fails to be fully voluntary.

What, then, distinguishes voluntary actions from involuntary or incompletely voluntary acts? The answer, according to Feinberg, is that an action that is caused by some external source or even inner compulsion may yet be voluntary, if it is what the agent would have done anyway. He uses the example of a man attached to a device that moves him according to the direction of the wind. The argument goes that the actions of this man are not his own; they are rather that of the machine – he does not move voluntarily. Feinberg's response is the following:

> "The reply to this argument, it seems to me, is that it does make a good deal of difference not into *what* we are plugged, but rather *how* we are plugged in. If the determining influences are filtered through our own network of predispositions, expectations, purposes, and values, if our own threshold requirements are carefully

observed, if there is no jarring and abrupt change in the course of our natural bent, then it seems to me to do no violence to common sense for us to claim the act as our own, even though its causal initiation be located in the external world. In short, the more like an easy triggering of a natural disposition an external cause is, the less difficulty there is in treating its effect as a voluntary action" (1970, 172).

Locke – makes the point that voluntariness does not entail avoidability: "it does not follow from the description of an act as voluntary (unconstrained, informed, and deliberate) that an agent could have done otherwise" (Feinberg (citing Lehrer) 185). If the shackled man's movements accord with his dispositions then they are voluntary, regardless of the fact that he could not have done otherwise.

On this view, the similarity between involuntary and non-voluntary acts need not be merely metaphorical. All the actions along the continuum are *caused* by something internal or external to the agent. But voluntary acts fall into a special category because they are in accordance with the predispositions, expectations, and so on, of the agent. As such, it is easy to understand why paternalistic interference in acts at the non-voluntary end of the continuum is justified. Since they are not in accordance with the agent's dispositions, they need not be regarded as expressions of sovereignty or self-determination.

Harry Frankfurt famously makes the point about voluntary action in 'Freedom of the Will and the Concept of a Person', where he argues that "The enjoyment of a free will means the satisfaction of certain desires – desires of the second or of higher orders – whereas its absence means their frustration" (Frankfurt 17). We are free, according to Frankfurt, when we are free to want what we want to want. Frankfurt's example of the unwilling addict can clarify how this way of understanding voluntariness (or freedom) helps to resolve the problems raised in the previous section. Recall the Van Gogh example: Van Gogh's illness causes him to do two things, mutilate his ear and paint beautiful pictures. For the latter we wish to praise him, and find it absurd to consider a paternalistic intervention. For the former, however, we are likely to excuse him, and if possible paternalistically intervene to prevent the damage.

The unwilling addict has two first-order desires. He wants to take the drug he is addicted to, but he also wants to abstain. Frankfurt writes:

"Both desires are his, to be sure; and whether he finally takes the drug or finally succeeds in refraining from taking it, he acts to satisfy what is in a literal sense his

> own desire. In either case he does something he himself wants to do, and he does it not because of some external influence whose aim happens to coincide with his own but because of his desire to do it. The unwilling addict identifies himself, however, through the formation of a second-order volition, with one rather than the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. It is in virtue of this identification and withdrawal, accomplished through the formation of a second-order volition, that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it" (Frankfurt 13).

Similarly, if Van Gogh identifies himself with his painting – if he has a second-order desire to paint, or if, in Feinberg's terms, painting fits his predispositions, expectations, and so on – then his painting is free or voluntary. In contrast, if we assume that he does not have a second-order desire to mutilate his own ear, the act is not voluntary: its causation by his illness is not tempered by its being in accordance with his second-order desires.

## 2.3 Freedom and determinism

"The notion that necessity does not inevitably undermine autonomy is familiar and widely accepted," writes Frankfurt. Moreover, "necessity is not only compatible with autonomy; it is in certain respects essential to it. There must be limits to our freedom if we are to have sufficient personal reality to exercise genuine autonomy at all. What has no boundaries has no shape" (Frankfurt 1988 ix). With this comment Frankfurt aims to bolster the type of response he gives to the problem raised above. But this quote also illustrates that the Frankfurt-style response does not merely allow us to evaluatively distinguish between two choices made under the same mental conditions. In postulating a notion of voluntariness that is independent of causal features, Frankfurt also provides a response to the traditional problem of scepticism about free agency. Frankfurt's account of voluntary action is compatibilist in arguing that there can be free will regardless of whether human actions are internally or externally caused.

The question of whether we can be free if determinism is true has been traditionally linked to that of whether we can be morally responsible if determinism is true. This is largely due to

the common idea that freedom is a necessary condition of moral responsibility – so if we are not free, we cannot be morally responsible. But our concern with freedom arises from the question of whether paternalistic interference in any action might be justified, rather than whether the agents are morally responsible for the action. For these reasons we will focus on the question of free will, rather than that of moral responsibility.

The puzzle about free will is a familiar one. Free will is argued to be incompatible with determinism. (Free will is also argued to be incompatible with indeterminism, though this argument will not concern us here – it is commonly held that even if the world is indeterministic at the quantum level, at the level of human behaviour it is for practical purposes deterministic.) A deterministic system can be characterised as one in which for any two propositions $S_1$ and $S_2$ expressing the state of the world at times $t_1$ and $t_2$ respectively, $S_1$ and the laws of physics (L) entail $S_2$: $(S_1 \& L) \rightarrow S_2$. If determinism is true and we could choose between courses of action (say, if at $t_2$ we could choose to raise one arm rather than the other), then by *modus tollens* we could falsify either the past or laws of physics. Assuming that we can't falsify the past or the laws of physics, we cannot choose between courses of action (Van Inwagen, 47-48).

## 2.4 The hierarchical response

Frankfurt's and Feinberg's responses to this puzzle have come to be known as 'mesh' compatibilist responses. What is required for freedom, on these views, is an appropriate sort of mesh between "an agent's choices or actions and her other actional constituents like desires and preferences" (Haji 210). Frankfurt's theory, at least, is of the hierarchical variety: the appropriate mesh is taken to be that between higher order and lower order volitions and desires (Haji 210-211). So, as indicated earlier, freedom does not require the Van Inwagen condition that we 'could have done otherwise'- thereby requiring us to falsify the past or the laws of physics. Freedom consists in the meshing of one's higher- and lower-order volitions.

Haji points that a number of problems have been raised over the hierarchical models of freedom (212-213). Two of these are particularly compelling. The first is that, on hierarchical models, a person is identified with some set of her higher order desires or preferences. So, for example, the unwilling addict is identified with his second-order preference not to take the drug to which he is addicted. This is what enables him coherently to say that he takes the

drug against his will. But the division of preferences and desires into first- and higher-order seems to be one of theoretical convenience rather than metaphysical significance. All of the desires and preferences – whether they be directed at the external world or at other desires and preferences – are equally the desires and preferences of that person. Agreeing to refer to preferences with different directions as being of different orders, should not permit a sleight of hand that accords preferences of a higher order a higher status.

The second problem for the hierarchical response to the problem of free will (or autonomy, as we might want to conceive of it) is one of regress. On the hierarchical model, a person's will is free is their second-order desires are effective – that is, if they are able to want (at the first-order) what they want (at the second-order) to want. Effective first-order desires arising from compulsions, addictions and so on, that are not validated at the second-order, are then taken to be indicators of an unfree will. But what guarantees that second-order desires are not the result of compulsions or addictions? Suppose I am told as a child to always be generous. Being an earnest type, I confess to my mother each time I act ungenerously, and she withholds my dessert. I am effectively conditioned to want (at the second-order) to want (at the first order) to give to others. It seems that Frankfurt must say that this second-order preference of mine can be an expression of free agency, provided it is validated at the *third* order. I must prefer to have been conditioned in such a way as to always want to want to give to others.

However, the problem can be raised at any order of the agent's preferences. Couldn't my third-order desire to want to want to want to give to others, be as alien to me as the unwilling addict's first-order desire to take the drug? Frankfurt's response is this:

> "It is possible to terminate such a series of acts [of identification at higher and higher orders] without cutting it off arbitrarily. When a person identifies himself *decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders. Consider a person who, without reservation or conflict, wants to be motivated by the desire to concentrate on his work. The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of desires or volitions of higher orders.... The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked" (Frankfurt 1971, 16).

At this point we may feel the value objection kick in again – why does a certain level of preference provide the decisive indicator of a person's wants? Why does identification at one level trump identification at another level? And how realistic is it to put forward a theory which attempts to impose a rigid hierarchical structure on the grasshopper thoughts that make up most human minds? This last objection seems to me most telling against Frankfurt's hierarchical model: it fails on simple credibility.

## 2.5 Gathering threads

The problem we had was how to distinguish voluntary from non-voluntary actions, given the problems of scepticism about free will and the difficulty of distinguishing between two choices made under apparently the same mental conditions. The worry is that if anything is non-voluntary, everything is: all our actions and 'choices' (not just those of the addict or the person swept down the river) are determined by factors beyond our control. Feinberg hints at a response to these problems in 'Causing Voluntary Action'. It is clear that he believes that not all action is non-voluntary, and that even caused action can be voluntary in the relevant sense. The relevant sense for us is that there must be some actions that can be assessed as voluntary enough to be immune from paternalistic interference.

Of course this is not a problem that only arises on Feinberg's model. Wherever autonomy or self-determination, or freedom, or sovereignty is touted as a value, the metaphysical problem of freedom of the will becomes an issue. Why make such a fuss about the right to self-determination, if we are deluded that we are self-determining in the first place? But Feinberg's analysis of when paternalism is justified and when it is not, seems particularly prone to the problem. Because Feinberg uses a comparison with *in*voluntary action to explain why it is justifiable to interfere in *non*-voluntary action, other features of involuntary action are naturally assimilated into non-voluntary action. Non-voluntary action is naturally taken to be no more of an expression of free agency than is involuntary action. But non-voluntary action is on a continuum with voluntary action (as Feinberg readily admits). Factors such as ignorance and internal compulsion come in degrees, and sufficiently voluntary is therefore only a short step from substantially non-voluntary.

Feinberg's comments in 'Causing Voluntary Action' could be understood as advocating the sort of hierarchical model of the free will that Harry Frankfurt supports. On Frankfurt's

model, a choice would be voluntary when there is a decisive commitment to it at a higher order of preference. Non-voluntary choices are still choices of that person, but we can make sense of why someone might say that their choice was alien to them, why they acted *against their will*. We can make sense of the comparison Feinberg makes between involuntary and non-voluntary choices. Because non-voluntary choices are not themselves *chosen*, non-voluntary choices *are* importantly like being swept down a river. On this model, the similarity might be strong enough to ground the moral distinction that Feinberg makes – that interference in non-voluntary choices is justified, where interference in voluntary choices is not.

But we have seen that there are major problems with this view of the distinction between voluntary and non-voluntary choices. There is no real distinction between desires and preferences of a lower order, and those of a higher order. Moreover, we encounter a regress problem when we try to decide which level of the person's will makes the final decision about what they want and who they are. Finally, Frankfurt's model is structurally dubious.

## 2.6 An alternative explanation of voluntariness

Feinberg's analysis of justified and unjustified paternalism needs some sort of compatibilist underpinning, to explain why not every action is not non-voluntary. But the hierarchical model is not the only available alternative. One other compelling response to the problem of how any action can be free, is Daniel Dennett's compatibilism.

Dennett's compatibilism is based on a thoroughgoing naturalistic perspective. His argument turns on the complexity of evolved life – human life in particular – and the way this complexity demands conceptualisation at more than one level. At the physical level, events may indeed be determined, such that "there is at any instant exactly one physically possible future" (Van Inwagen quoted in Dennett 2003, 25). But the physical level is not the only one at which we can conceptualise life or events. In our world, there is also the design level. The (probably deterministic) processes of evolution in our world have created over time entities that can be thought of as if they were *designed* by the hand of evolution. The *design level* "has its own language, a transparent foreshortening of the tedious descriptions one could give at the *physical level*" (Dennett 2003, 39). The ontology of the design level, as Dennett points out, will be significantly different from the ontology of the physical level. At the most basic

physical level entities can clearly be seen to obey the laws of the deterministic universe. But as more basic elements are added and the corpus of laws expanded, "simple entities can evolve that are capable of avoiding harm and reproducing themselves" (22).

This evolution of complexity can be illustrated in simple software toys such as the game 'Life', in which complex, apparently *designed* systems evolve from very basic components with only two possible states, and simple deterministic laws. 'Life' at the design level suddenly starts to appear purposeful: clusters of elements 'glide' across the screen, or 'eat' each other in order to 'survive'. In Dennett's view our actual, physical world is analogous to the 'Life' world. Though the basic physical elements are indeed basic, unintelligent and governed by deterministic laws, the sheer number of basic elements and deterministic laws creates an overwhelming complexity that demands conceptualisation at more than one level.

Dennett argues that the concept of inevitability – traditionally linked to determinism – properly "belongs at the design level, not the physical level." (22) The concept of inevitability is grounded in that of *evitability*, the possibility of avoidance. Without the possibility of avoidance – of harm, for instance – we could never grasp the *im*possibility of avoidance that is inevitability. But it is at the design level that there is *action* rather than mere occurrence. (43) In our complex world, some creatures have been 'designed' to be able to avoid harm. Since some harms are avoidable by creatures at the design level, some harms are evitable. Evitability and its dependent opposite, inevitability, are evolved concepts of the design level. Not everything at the design level is inevitable, and so the traditional slip from determinism to inevitability is blocked. In Dennett's view, it is a failure to appreciate the complexity of the natural, evolved world, including the personal and cultural worlds of humans, that has given us the mistaken belief that some of our concepts must clash.

Dennett argues that "Determinism doesn't imply that whatever we do, we could not have done otherwise" (95). This misconception about determinism arises from a misunderstanding of the concept of possibility. Dennett casts the concepts of necessity, possibility and causation in terms of possible worlds. A necessity is true in all possible worlds, and, according to Dennett, a possibility is whatever is not necessarily not the case. But when we canvass the possible worlds to determine whether something is necessary or possible, we can restrict the scope of our search more or less, depending on the context (66-67). In particular, when we assess counterfactual claims such as 'If you had tripped Arthur, he would have

fallen', we pick out a set of worlds substantially similar to our own in which you tripped Arthur, and see whether he fell in all, some, or none of them.

To simplify Dennett's position, consider some action and its consequence in this world – I struck a key, and a letter appeared on my monitor. Could it have been the case that the letter did not appear? From the perspective of this world, at the instant I hit the key, there is only one *physically* possible world, given determinism. (This is the definition of determinism.) But, had the world been ever so slightly different up to that point (and it might well have been) there is no confusion in saying that the letter need not have appeared. It is this variability in the scope of what we take as the basis for deciding what could have been, that leads to confusion about whether we could have done otherwise (70; 75-77): "the truth or falsity of determinism should not affect our belief that certain unrealized events were nevertheless "possible", in an *important everyday sense* of the word" (77).

To some, this may seem like cold comfort – certainly, I could have done otherwise, but only if I lived in an entirely different possible world! This intractably sceptical position would say that it *is* what is physically possible that concerns us in these cases: it is important that we could not, physically, have done otherwise. But let's grant Dennett his construal of possibility for the moment and see how he believes freedom evolves.

Dennett sees freedom as a human construction, like music or romantic love – but like music and love, the reality of freedom cannot long be denied. Though Dennett's analysis requires greater conceptual tinkering that Frankfurt's – of our concepts of possibility and inevitability, for instance – it is still a compatibilist response to the problem of freedom given determinism. We are free when we *could* have done otherwise, with the 'could' opening up a class of actions that are consistent with our 'design', rather than a class of physical possibility (since our actual world is the only physically possible one, given some previous state and deterministic laws). Unlike Frankfurt's model, Dennett's is based in a realistic and complex picture of the human mind.

## 2.7 The nagging doubt

If freedom is a human creation like music or love, then it is subject to human destruction. Dennett argues that his variety of freedom – based on scientific understanding of the processes of evolution rather than "false myths about human nature" (Dennett 2003, 287) is both more worth defending and more robust than older conceptions of freedom.

But there is a question of 'creeping exculpation'.

> "We now uncontroversially exculpate or mitigate in many cases that our ancestors would have dealt with much more harshly. Is this progress or are we all going soft on sin? To the fearful, this revision looks like erosion, and to the hopeful it looks like growing enlightenment, but there is also a neutral perspective from which to view the process. It looks to an evolutionist like a rolling equilibrium, never quiet for long, the relatively stable outcome of a series of innovations and counter-innovations, adjustments and meta-adjustments, an arms race that generates at least one sort of progress: growing self-knowledge, growing sophistication about who we are and what we are, and what we can and cannot do. And from this self-understanding, we fashion and re-fashion our conclusions about what we ought to do" (Dennett 2003, 290).

We should not be frightened of the impact of a greater understanding of our place in the natural world, according to Dennett. Rather, we should explore these implications to the full and decide what we are to do about them.

Dennett's view is that our institutions of moral praise and blame should not be swept aside by this new understanding – this is unsurprising given he has a picture of the evolutionary importance of these institutions. (Though of course some features of human anatomy and society that were once very useful are no longer so.) But, to use one of Dennett's catchphrases, "If you make yourself really small, you can externalize virtually everything." If I understand myself as a product of the shaping forces of evolution, and my genetic code, and my upbringing, then it seems I need no longer take responsibility for my actions.

Dennett's response to this problem involves what he sees as a shift in our thinking about the ethical and the metaphysical.

> "The key shift in perspective that will enable this is an inversion described by Stephen White in *The Unity of the Self* [...]. Don't try to use metaphysics to ground ethics, he argues; put it the other way around: Use ethics to fix what we should mean by our "metaphysical" criterion. First, show how there can be an internal justification for some agent acquiescing in his own punishment – saying, in effect, "Thanks, I needed that!" – and then use that understanding to anchor and support a reading of our pivotal phrase, *could have done otherwise*: "An agent could have done other than he or she did just in case the ascription of responsibility and blame to that agent for the action in question is justified" (p.236). In other words, the fact that free will *is* worth wanting can be used to anchor our conception of free will in a way metaphysical myths fail to do" (Dennett 2003, 297).

What could possibly make someone acquiesce in her own punishment? According to Dennett, it is that she values being treated as a free agent in other contexts. "Since there will always be strong temptations to make yourself really small, to externalize the causes of your actions and deny responsibility, the way to counteract these is to make people an offer they can't refuse: If you want to be free, you must *take* responsibility" (292).

This point was already mentioned in the context of the Van Gogh problem: freedom was seen to be necessary for *identity*, because freedom is required for responsibility, and responsibility for identity. But the Van Gogh problem can be raised again here to make a different point. A possible problem for Dennett is that many people will want to be treated as autonomous agents in situations where it is to their advantage but not otherwise. The ascription of blame will not be justified unless the person acquiesces *for that action.*

The Van Gogh example shows that it is important to distinguish between acts made under similar mental conditions, to allow that responsibility can be held for the one but not the other. But given this possibility, it seems that in the context of *moral* responsibility we might pick and choose what actions we want to be held responsible for. Perhaps Dennett can address this problem by giving special consideration to the chronically mentally ill. Or maybe there are simply fewer difficulties in granting general responsibility than moral responsibility. We will leave these problems of the practical stance for the moment, to bring the discussion back to paternalism.

## 2.8 The flip side: creeping paternalism

Dennett admits that we are likely to exculpate in more and more cases, given a greater understanding of the mind. But this allowance of diminished responsibility in a broader range of cases invites a 'creeping paternalism' as well: in cases where responsibility may legitimately be denied by the agent, paternalistic intervention becomes a live option. But to deny responsibility is to deny the self – and who wants to do that? Thus what holds the line against creeping exculpation should also hold the line against a creeping paternalism – *if* the practical stance is workable at all.

How is this not like the compatibilist solutions of old? Well, it makes explicit what was already there – that we're basing the idea of freedom on our moral intuitions rather than the other way around. Whether this is legitimate can be put to a practical test, and at least is an idea worth exploring.

## 2.9 Conclusion

We started this chapter with the question of whether Feinberg's distinction between voluntary and non-voluntary choices can do the work that he wants it to do in his theory of justified versus unjustified paternalism. What we think of as non-voluntary choices – those made when we are under mental pressure, or irrationality, or significant ignorance – are traditionally linked by metaphor to completely involuntary actions to which the concept of choice does not apply at all. Feinberg makes use of this metaphorical linkage to argue that non-voluntary choices, like involuntary actions, are not properly thought of as expressions of self-determination or sovereignty.

But the link is only metaphorical. Non-voluntary choices and involuntary actions are more dissimilar than they are alike. In particular, whether an act is held to be non-voluntary or voluntary depends at least in part on how it is viewed by the person who makes it, and those who observe it or talk about it. Involuntariness (though not entirely clear-cut) seems to be more objective. The standard solution to this problem has been a hierarchical model of an individual's choices, such as that proposed by Harry Frankfurt. On this model, the voluntary choice is the one that is chosen, while the involuntary choice is the one the individual would prefer not to make. This is how we are to make sense of how non-voluntary choices are

importantly like involuntary actions. Because people would rather not make certain choices, these can be seen as somehow external to them.

But apart from the nagging sceptical worry, there are problems with Frankfurt's view, the most serious of which are that it requires an unrealistic division of the self into levels of choice, and places unwarranted value on the 'higher' levels. *Identity* does not reside at one or the other level of preference in the human mind, though consciousness may play a large part in stringing it all together for us. *Identification*, therefore, cannot occur at any single level.

Dennett's compatibilism is based on a complex and thoroughgoing naturalism. Which choices are made freely – when we *could* have done otherwise – will be something we figure out on the basis of the overall responsibility we want to take, on the size of self we want. Thus freedom becomes dependent on our practises of responsibility-attribution. And since paternalistic action and law are instances of responsibility-attribution, the freedom we attribute to individuals will depend partly on our practices of paternalism.

## Chapter 3

### 3.0 Introduction

It is a presumption of the law, medicine and liberal theory (including Mill's *On Liberty*) that there are cases where we are completely justified in acting paternalistically, as we have defined paternalism above. That is, we are justified in usurping some choices for the purpose of promoting the good of the person concerned. The easiest cases of where paternalism seems justified are those of self-harming choices made by children and people suffering from dementia and other debilitating mental illnesses. Feinberg attempted to ground this 'soft' paternalism by arguing that it does not violate the fundamental right of sovereignty because (along with choices made from physical pressure and coercion) the choices are non-voluntary. But when we asked exactly what distinguished voluntary choices from the non-voluntary, we came across some difficulties. At the practical level, the Van Gogh problem illustrates the difficulty of distinguishing between voluntary and non-voluntary choices made by the same person in the same state. Then there is the underlying metaphysical problem of whether there can be voluntary choices at all.

In this chapter I begin to explore Dennett's idea that metaphysical freedom be based on actual ascriptions of responsibility. Dennett argues that what is classed as a voluntary action should be based on what, in practise, we take and are given responsibility for. In the practical context of paternalism, such ascriptions of responsibility are made through determinations of *competence* in medicine, either for application in medicine or for legal reasons. A person is deemed responsible for a criminal action, to take the relevant example, only if she is deemed a competent decision-maker. Thus the competence determinations common in medical practise and law will set off a class of voluntary actions from those that are non-voluntary. The incompetent person is denied responsibility for her choices. These choices will therefore be classed as non-voluntary, and may be paternalistically usurped, according to Feinberg's model.

## 3.1 Competence

The standard practical view on paternalism is that we are at least justified in denying choices made by the *incompetent* person. The Millian intuitions about when paternalism is legitimate – in cases of derangement, ignorance, childhood, and so on – are formalised in practical guidelines for determinations of competence. They are subsequently enforced through law that takes competence as setting the boundary of criminal responsibility. In *Deciding for Others*, Allen Buchanan and Dan Brock set out detailed guidelines for determining the competence of individuals to make decisions for themselves. Their guidelines constitute the practical side of the equation that Dennett wants to balance.

Incompetence, Buchanan and Brock argue, is decision-relative – that is, a person may be competent to make one decision but not another. Or a person's competence for a particular decision may be affected by her surroundings, by who explains the alternatives to her, by the current state of her illness, and so on. For these reasons it is not possible to make an overall, one-off assessment of a person's decision-making competence (Buchanan and Brock (B&B) 18-20). It is the decision-relativity of competence determinations that is thought to solve problems like the Van Gogh problem. For if competence is a function of particular decisions rather than simply the mental state of the individual concerned, Van Gogh can be held competent to paint but not to self-mutilate.

Buchanan and Brock distinguish three crucial elements of decision-making competence. The competent person

1)can understand and communicate,

2)can reason and deliberate and

3)has a set of values or a conception of the good.

Understanding and the capacity to communicate are required for the person to be adequately informed as to what alternatives they have, and to express their eventual choice. Interestingly, Buchanan and Brock set quite a high standard for understanding:

> "Understanding is not a merely formal or abstract process, but also requires the ability to appreciate the nature and meaning of potential alternatives – what it would be like and "feel" like to be in possible future states and to undergo various experiences – and to integrate this appreciation into one's decision making" (B&B 24).

The criterion of being able to understand the impact of one's decisions is another reason why competence will be decision-relative. For instance, a child might be allowed to decide what sweet she wants, since she has the capacity to understand the possible consequences of this decision (eating what she would like to eat, not eating something she doesn't want), but not allowed to eat sweets all the time, since she can't fully comprehend the long-term possibilities that hinge on her decision.

The second criterion – the capacity for reasoning and deliberation – is required so that the process of decision-making can take place. "Reasoning and deliberation require capacities to draw inferences about the consequences of making a certain choice and to compare alternative outcomes based on how they further one's good or promote one's ends" (25). Reasoning and deliberation seem to be the calculative component of choice, where understanding is the ability to put oneself in the position of having made the choice, and seeing how one would feel. And finally, the competent decision maker must have a conception of the good or set of values by which to judge the various outcomes open to her, and this conception must be "at least minimally consistent, stable, and affirmed as his or her own" (25).

Buchanan and Brock stress that decision-making competence is a threshold concept, not a comparative one: "in the legal process in particular, it is important to resist the notion that persons can be determined to be more or less competent, or competent to some degree, to make a particular decision… [The function of the determination of competence is] first and foremost, to sort persons into two classes: (1) those whose voluntary decisions (about their health care, financial affairs, and so on) must be respected by others and accepted as binding, and (2) those whose decisions, even if uncoerced, will be set aside and for whom others will act as surrogate decision-makers" (B&B 27). (By 'voluntary' Buchanan and Brock simply mean unforced and uncoerced.) Determinations of responsibility for medicine and law need to be clear cut. Competence, therefore, cannot be held by degrees.

However, all of the components of competence may be held by degrees. Understanding and communication can be impaired to a greater or lesser extent. A person learning about her medical condition in a language not native to her has some difficulty in understanding and communicating, while the unconscious patient has no ability to understand or communicate at all. Reasoning and deliberation abilities are even more variable and difficult to assess –

human intelligence is an active area of research in which the only certainty is that there are no clear boundaries. Finally, the set of values on which decisions are to be based may be more or less coherent and well-defined. It is therefore clear that the overall determination of competence is based on some kind of (implicitly or explicitly) *stipulated* threshold: no boundary of competence exists in nature.

### 3.2 Setting standards of competence

Buchanan and Brock differ from Feinberg in arguing that there is a *positive* case for paternalism. They agree that respect for a person's self-determination is one of the most important values at stake, but the paternalist is positively justified because her action promotes and protects the good of the individual whose choices are usurped. For Buchanan and Brock, self-determination and a person's good can conflict, and the two must be balanced: self-determination is not overriding. If a patient is judged incompetent to make a particular decision then her self-determination may be sacrificed in favour of her good. But the decision as to whether a patient's self-determination may be sacrificed to further her other interests will depend on features of the particular situation. "There is no uniquely "correct" answer to the relative weight that should be assigned to these two values, and in any event it is simply a fact that different persons do assign them significantly different weight" (B&B 41).

In practise, what this means is that the threshold level of competence is stipulated in response to a relative assignment of value to the self-determination and overall good of the person concerned. If self-determination is found to be weightier in one context, then the threshold level of competence may be set lower. But if the person's good is taken to have more value, the threshold of competence may be set higher, and therefore be more difficult to meet. It is more likely in this situation that paternalistic interference will be warranted.

In what situations might a person be judged incompetent on the basis of an assessment of these values rather than an objective analysis of their mental state according to the criteria outlined above? In borderline cases, factors such as the riskiness of the decision come into play. If a person whose competence is questionable wants to do something that will have a seriously negative impact on their well-being, either in the short or long term, then concern

for her good might outweigh concern for her well-being, and she will be judged incompetent on this basis.

Here we can see clearly that this theory operates at the practical rather than the normative theoretical or metaphysical level. For unless they are willing to defend an ethical particularism (and Buchanan and Brock do not indicate that this is their aim), philosophers *will* want to assign reasonably definitive relative weights to the values, and will not be overly concerned by the fact that people do assign them different weights in different circumstances. The practical inconsistency of humans need not be incorporated into a theory of the virtues. This fact about how people actually weigh self-determination and the overall good seems likely to lead to confusion in this context anyway, since it creates an ambiguity about whether we should take the patient's or the doctor's assignment of value as conclusive. What is important for the moment, though, is that Buchanan and Brock take the two values of self-determination and the good of the person to require balancing before a judgement of whether paternalism is warranted can be made.

In fact, Buchanan and Brock argue that Feinberg's attempt to rest his theory of justified paternalism on the single value of sovereignty is incoherent. We have seen that for Feinberg, where self-fulfillment and sovereignty conflict, sovereignty always dominates (remember that non-voluntary choices are not expressions of sovereignty). "Feinberg rejects perfect voluntariness as an impossibly high standard, so the question is how voluntary is "truly voluntary" or voluntary *enough* to be protected from paternalistic interferences...Feinberg offers some rules of thumb for determining the appropriate level of voluntariness, two of which are that the more risky the conduct, and the more irreversible the risked harm, the greater the degree of voluntariness that should be required if the conduct is to be permitted" (B&B 43-44). But Buchanan and Brock argue that Feinberg's voluntariness assessment involves a trade-off between self-determination and the person's good. The fact that the degree of potential harm and the riskiness of the venture make some input into the determination of voluntariness, on Feinberg's model, indicates that self-determination is not the only value at stake. By introducing these factors Feinberg is effectively doing exactly what Buchanan and Brock do: weighing the potential damage to the person's interests against their right to be self-determining.

Thus Buchanan and Brock see little conflict between their own view and Feinberg's – both balance self-determination and the individual's good in the crucial determinations of which choices should be allowed to be carried through without intervention. The more stringently we set standards of competence (or voluntariness) the more we opt for promoting the person's good at the expense of self-determination. The lower our standard of competence, the more our concern for respecting self-determination at the risk of damaging overall interests.

Buchanan and Brock write:

> "The proper standard of competence must be chosen; it cannot be discovered. There is no reason to believe that there is one and only one optimal trade-off between the competing values of well-being and self-determination, nor, hence, any one uniquely correct level of capacity at which to set the threshold of competence – even for a particular decision under specified circumstances. In this sense, setting a standard for competence is a value choice, not solely a scientific or factual matter" (B&B 47).

According to Buchanan and Brock's model, Van Gogh's two decisions can be distinguished by their riskiness. When he wished to paint, Van Gogh's decision does not pose any threat to his future well-being – thus he can safely be judged competent to paint. But when he wants to self-mutilate, we might take his good to outweigh his self-determination, and judge him incompetent to make the decision. The threshold for competence becomes more stringent as the desired behaviour becomes riskier.

One might legitimately wonder how we got from the threefold, reasonably objective criteria for determining competence to the claim that competence (and voluntariness, by Buchanan and Brock's argument) is a *value* judgement. For the threefold criteria it would seem that competence is a judgement that can be made purely by psychological assessment of the patient, rather than any weighing of values. But on Buchanan and Brock's model, competence ends up having little to do with mental capacities and more to do with the evaluation of the decision made by those who are in the position to set the threshold for competence.

But the idea that a person's competence can be judged in terms of relative valuations of her good and her self-determination is absurd: her competence is a feature emergent on her mental capacities, not on the weighing of values by some outside observer. The riskiness of

the Van Gogh's choice has nothing constitutive to do with his mental competence (though it might be taken as evidence for a mental problem). If the riskiness of decisions were taken as the deciding factor in determining incompetence, most of the world would be uninhabited and space unexplored, as well as countless scientific discoveries unmade. The fact is that reasonable people can calculate that an extremely dangerous activity is worth the risk to them, though others may be unable to comprehend the potential benefits.

The problem for Buchanan and Brock's theory arose when we came to the point of stipulating a threshold of competence. For, as we saw, all the elements of competence can be held by degree, and there seems to be no natural way of separating the competent from the incompetent individual. However, this does not mean the competence judgement must become a value judgement. There are legitimate ways of stipulating a threshold of competence that do not require weighing self-determination against the person's good in the first instance.

The Van Gogh problem has an alternative solution to that mentioned above. Though competence determinations need not be directly risk-related, they can be risk-related in an indirect manner: it may take more mental capacity to comprehend the consequences of a risky decision than a safe one. We saw that a child might be competent to decide what to eat for a particular dessert but not to eat sweets for every meal, because she can comprehend the possible outcomes in one decision-making situation but not the other. Likewise, Van Gogh may be competent to make the decision to paint because the outcomes of painting are less complex and require less mental capacity to comprehend than the outcomes of self-mutilation. If Van Gogh does not have the capacity to understand the outcome of self-mutilation – what it will be like to be mutilated – then he may be judged incompetent to make the decision to self-mutilate.

## 3.3 The story so far

Joel Feinberg sets out the limits of justified paternalism in terms of voluntary and non-voluntary action. Paternalism is justified when the choices at stake are substantially non-voluntary, because those choices are not to be regraded as expressions of sovereignty. Feinberg's account, like other ethical theories at the applied level, rests on a distinction between classes of actions that seems eventually to require an answer to the problem of

whether any actions can be free at all. Daniel Dennett argues that many of our actions are freely chosen in a strong sense, but that we are looking for the basis of that freedom from the wrong perspective. Rather than trying to ground the ethical on the metaphysical, we should be using our ethical practises to ground our metaphysics. What we take and are given responsibility for, according to Dennett, is what we do voluntarily.

One class of actions that we are *not* given responsibility for, are those done when we are judged to be incompetent as decision-makers. Young children are not held to be responsible for many decisions because they do not have sufficient experience or reasoning ability to be able to comprehend the consequences of their actions. The same goes for psychotic and senile individuals. Since, in the practical context, responsibility is withheld from these individuals, it follows (on Dennett's model) that their actions are not undertaken voluntarily. What counts as incompetence? Buchanan and Brock set out the requirements for competence in terms of mental capacities. But mental capacities come by degrees, while competence judgements are either positive or negative. Therefore a threshold for competence must be stipulated.

### 3.4 Begging the question in competence judgements

Buchanan and Brock argue that the threshold for decision-making competence should be made by referral to the values of self-determination and the overall good of the individual whose competence is in question. But if these are the deciding factors in competence judgements, why bother with independent tests of mental capacities at all? Why not just decide whether we value their self-determination or their overall good more highly, and let that be the decider of their competence: they are incompetent when we want to help them against their wishes and competent otherwise?

To put the point more precisely, when we are making a decision about someone's decision-making competence, we are trying to determine whether it is legitimate or not to interfere in their choices. Using *our* evaluation of their self-determination versus their overall good as the *basis* for deciding whether they are competent or not, thus begs the question as to whether it is legitimate to interfere in their choices. Our competence determination is supposed to be the decider, but we have already decided when we weigh overall good more heavily than self-determination, and set the competence threshold too high for that individual to meet.

Begging the question of the legitimacy of interference in competence determinations can clearly lead to violations of the right to self-determination – in effect, it lets hard paternalism in through the back door. If competence – and therefore voluntariness – is decided by reference to the evaluator's judgement of the respective value of self-determination in this particular case, the evaluator is free to rank self-determination below overall good whenever the choice in question is not one they support. This is particularly problematic when the desired behaviour is extremely risky or damaging. As I argued earlier, the fact that a choice is risky has nothing to do with whether a person is competent to make the choice, except insofar as greater mental capacities may be required to assess the possible outcomes of the choice. Standards of competence should not be automatically higher for risky choices. But if the threshold for competence is set with respect to a valuation of a person's good, it will turn out to be risk-related in just this way: it will be legitimate to stop people with the mental capacities of a reasonable adult from undertaking extremely risky actions. This point will become clearer as we explore a case of possible paternalism in the law.

But let us quickly look at how this relates back to the practical stance on the metaphysical problem of voluntariness advocated by Dennett. Recall that, according to Dennett, the class of voluntary actions should be delineated by what we take and are given responsibility for in practical situations. In law and medicine, judgements of incompetence separate a class of decisions that we are not held responsible for, from the rest. Decisions made when we are incompetent may be paternalistically usurped, and nor are we blamed for them by the law. The actions resulting from decisions made by incompetents, then, will be a class of non-voluntary actions, effectively unfree.

But Dennett's practical stance conflicts with the risk-related model of incompetence of Buchanan and Brock. Firstly, the riskiness of a choice seems to have no more to do with freedom than it does with mental capability. It does not seem that we become less free the greater the danger of the activity we wish to undertake. The reasonable adult skydiver is no less free than the reasonable adult bookworm. But on Buchanan and Brock's model we would get precisely this result. Since we are more likely to be judged incompetent when the choice is a risky one, we are given less responsibility for risky choices and are therefore less free when we make them.

## 3.5 Competence in the legal context

In the current law, consent is immaterial when an act occasions harm beyond a certain level. This was demonstrated in the case of *R* v *Brown*, in which five men were convicted of assault occasioning actual bodily harm and unlawful wounding on the basis of their participation in entirely consensual sadomasochistic activities. Though no complaints were made to the police, and the Crown's case was made on the evidence of video tapes members of the group had made, the conviction was upheld in the House of Lords. There were a number of strands of argument in the Lords' response to the appeal (including arguments from the immorality of the acts), but one at least was paternalistic: that persons simply should not be allowed to consent to harm beyond the level of 'transient and trifling' (*R* v *Brown* 75; 77-79; 88; 94).

There is a general principle in law that consensual acts are allowable. The restriction on this principle imposed by the House of Lords in the case mentioned above does turn out to be (at least partly) paternalistic, on the definition of legal paternalism with which this thesis began. In the next sections I explore another potential application of this restriction: cases of persons who wish to contract to have healthy limbs amputated. I start by determining whether the law against contracting to have such a harm performed upon one is paternalistic in the presumptively justified or unjustified sense, in terms of competence. On the practical model of competence, paternalism is presumptively justified when the person in question is incompetent as a decision-maker, and presumptively unjustified otherwise. I briefly rehearse the argument that it is not a sufficient justification for legal sanctions that they protect the good of *competent* individuals at the expense of their self-determination. If it is interpreted as an instance of presumptively justified paternalism, we must read this law as the claim that persons who wish to have a certain degree of harm inflicted upon them are to be presumed incompetent to make that decision.

The case of apotemnophilia – the desire to have a healthy limb amputated – may tempt us towards this view. We need not hold that the desire for such harm is constitutive of incompetence, merely that it is overwhelming evidence for incompetence. The law in effect stipulates a threshold for incompetence based on strongly valuing the good of those it affects over their self-determination in this regard. Those who wish to have harm beyond a certain level inflicted upon them, are deemed incompetent to make the decision on the basis of a stipulated threshold of competence that is higher than they can meet. But even in the case of

the desire to have a healthy limb amputated, this stipulation of incompetence can be called into question. The case will illustrate the problem of using evaluations of the weight of self-determination in a decision about competence. This problem will occur even if the individuals are assessed one by one rather than covered by the same law banning consent to serious harm.

We turn next to a general investigation of the proper response for the law to take in cases where the competence of decision-makers is questionable. As argued earlier, valuing the patient's good over their self-determination in stipulations of incompetence seems to be letting hard paternalism in the back door. To understand how to deal with cases of questionable competence we need first to understand the grounds for opposing paternalism involving competent individuals. This task turns out to present a number of philosophical challenges.

## 3.6 Apotemnophilia

In January 2000 the British press began to make much of the phenomenon of apotemnophilia. Dr Robert Smith, a surgeon at Falkirk and District Royal Infirmary in Scotland, had performed two amputations of healthy legs at the request of the patients, and with the support of the hospital's medical director and its chief executive. The patients were each assessed by two psychiatrists and a psychologist prior to surgery; they paid for the operations themselves, and Smith waived his personal fee. Both patients said they were very satisfied with the results of the amputations.

These newsworthy and somewhat sanitised cases of apotemnophilia are just the tip of the iceberg. In May 1998 a 79 year old man died of gangrene in a hotel room in Mexico, two days after a botched operation to remove his healthy left leg. The operation was performed by an ex-doctor who had been charged with incompetence and had his license revoked in 1977. In October 1999 a Milwaukee man severed his own arm with a home-made guillotine and threatened to sue the hospital and re-amputate his arm if it was reattached. On the premise that amputation can't be ethically denied if an injury would cause death without it, the website http://www.ampulove.com/ contains self-mutilation tips for 'wannabes' and gruesome photos of home surgery success stories.

John Hopkins psychologist John Money coined the term 'apotemnophilia' in 1977 (Elliott 2000) The suffix '-philia' might seem to place it in the category of sexual paraphilias, but it is not at all clear that all apotemnophiliacs are sexually motivated. Apotemnophilia has also been likened to body dysmorphic disorder and transsexuality: in both cases body image and identity seem to be of more importance than sexual gratification. (Elliott 2000) Apotemnophilia should be distinguished from acrotomophilia, a sexual attraction to amputees; acrotomophiliacs online are known as 'devotees'. It seems that apotemnophiliacs, like transgendered individuals, see themselves as being trapped in the wrong kind of body. The only thing that will satisfy them is changing their body to match their self-image. In the case of apotemnophiliacs, this means having a limb removed. If that means removing it themselves, many will go to that extreme.

Plainly, cases of apotemnophilia raise issues for a number of disciplines. What is of interest to us is the question of whether there is a case for legal intervention in medically assisted elective amputations such as those carried out by Dr Smith. Should they be allowed to have the limb amputated, or should this be prevented, for their own good? The case is an interesting one because of the level of harm the individuals wish to have performed upon them, and the paternalistic 'purity' of any intervention. Unlike tattooing, piercing, and many cosmetic surgeries, the harm involved is serious and permanent, but it is not as serious or permanent as euthanasia. It is therefore more difficult to make both the argument that surgical amputation of a limb is a private choice to be respected, *and* the argument that it is the product of a depressed state, and that the apotemnophiliacs are therefore incompetent to make the decision. By paternalistic 'purity' I mean that the motivation for any intervention in this choice would be largely paternalistic. Contrast the case of cigarette smoking. Though a policy of heavy taxation on cigarettes may have some basis in legal paternalism, there is a good argument for it from the burden on public health systems of the largely inevitable diseases of a large number of smokers. Apotemnophiliacs, however, are few, and the possible burden on the state of their disability is so negligible that it may be ignored. (As we accept the burden of those who choose to play dangerous sports.)

In current British law a person cannot consent to harm beyond a certain level of seriousness, and the amputation of a limb certainly falls beyond this threshold. In general, an act is permissible if it is consensual, but the current law places constraints on this principle of consensuality. "Starting from the premise that an assault necessarily entails the absence of

consent", Kell writes, "the general rule is that when the injury caused...by the consensual, dangerous conduct, reaches the level of 'bodily harm', as defined, the consent of the participants is rendered ineffective, and, consequently, incapable of preventing criminal liability attaching" (Kell 121). Bodily harm is defined as harm beyond the merely "transient and trifling". The law barring consent to harm beyond a certain point seems to be paternalistic: if the motivations for the law are as they seem, then it fulfils all our criteria for paternalistic state action. The state aims to bring it about that with respect to some state(s) of affairs which concerns Q's own good Q's choice or opportunity to choose is denied or diminished. A person who wishes to have harm beyond the level of transient and trifling performed upon them is not legally able to do so. The state's belief that this policy or law promotes the good of Q is the main reason for the implementation of that policy or law; we can, for the purposes of discussion, discount other reasons (such as public liability) the state may have for imposing such a restraint on the principle of consent. Finally, the state discounts any belief Q might have that its law or policy does not promote Q's good. Those who desire to have harm performed upon them clearly believe that it is in their own interests.

### 3.7 Paternalism and the choices of competent individuals

But what sort of paternalism is it? Presumptively unjustified paternalism on Buchanan and Brock's model involves the usurpation of the choices of *competent* adults, while justified paternalism involves interference in decisions made by people who are *incompetent* to make those decisions. So whether or not the law disallowing consent to harm beyond a certain level of seriousness is presumptively justified or unjustified according to their model depends on the competence of those affected by it.

The case against paternalistic interference in the choices of entirely competent individuals has been argued more eloquently and more persuasively than I can hope to match by liberal thinkers such as Mill. It is not the law's part to impose the state's conception of the good on competent adults at the expense of their self-determination. The point of Mill's Harm Principle is that it is dangerous to give the law power over the self-determination of an individual. This is not to say that the state can have no role in promoting as well as protecting the well-being of citizens. Mill writes that

> "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good,

> either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil, in case he does otherwise" (Mill 15).

Just as a person may entreat or reason with friends to try to get them to stop smoking or drinking, society can use methods other than force or coercion to try to improve the well-being of citizens. The least coercive and most effective means society has at its disposal is probably education. But the use of the law to attempt to promote the good of competent adults at the expense of their self-determination is unjustifiable. I will go into the liberal arguments against this variety of paternalism in more detail further along.

### 3.8 Paternalism and the choices of incompetent individuals

If the law barring consent to harm beyond a certain level cannot be justified when it is understood as an instance of paternalistic interference in competent choices, we might wish to interpret it as ranging over the choices of incompetents. The people affected by the law are those who wish to contract to have a fairly extreme harm caused to themselves; the choice at stake is the choice to enter into such a contract. For the law to be a genuine case of presumptively justified paternalism, it must be based on the assumption that all those affected by it are incompetent to make the particular decision to contract to have harm done to them. Given the fact that the law covers all sorts of cases where people contract to all sorts of different harms, the only possible grounding for the assumption of incompetence is that the very decision to contract to have harm caused to oneself beyond the level of bodily harm, is evidence for or constitutive of incompetence beyond the threshold level.

Alternatively, a law might be put into effect that only those judged competent by medical professionals be allowed to contract to have extreme harm such as the amputation of a limb performed upon them. In this case, the restrictive law would clearly apply only to those judged incompetent to make the decision. But on Buchanan and Brock's risk-related model, apotemnophiliacs will be likely to be judged incompetent simply because of the extreme and irreversible nature of the harm they wish to have performed. Thus contracting to have a healthy limb amputated is effectively ruled out in either case.

We might be tempted to the view that this restriction is warranted when we look at the case of apotemnophilia. The consequences of the loss of a limb are so dire that it may be thought that anyone who desires the amputation of a healthy limb is obviously incompetent to make the choice, and the legal restriction on this choice is justified. Not only will the individual have to face the pain, risk, and physical restrictions that result from a major surgery, but they will have difficulty with mobility and probably social acceptance for the rest of their life. Moreover, if we have trouble believing in the force of the apotemnophiliac's positive justification for surgery, then we will not accept that the decision has any merits to outweigh its negative impact.

But even in such an extreme case of the desire to self-harm as apotemnophilia, there are a number of considerations against this view. The apotemnophiliac desire for amputation is so strong that many sufferers will attempt to remove the limb themselves, or at least inflict such grave injury on themselves that amputation is the only option for the doctors to whom they turn. As well as the cases mentioned at the beginning of this chapter, there are reports of apotemnophiliacs who have attempted or seriously consider shooting their legs off, lying on railway tracks and attempting to induce gangrene in serious self-inflicted wounds. These are pragmatic reasons for failing to implement a justified paternalistic policy, as is the fact that there is no known cure for apotemnophilia. So far psychotherapy has not succeeded in diminishing the desire to have limbs amputated.

But those who have succeeded in having the limb amputated report great relief. Thus amputation, though it can hardly be seen as a 'cure' in the usual sense, is currently the best possible outcome for genuine apotemnophiliacs. The fact that apotemnophiliacs report satisfaction post-amputation tells against the view that, pre-surgery, they do not fully comprehend the consequences of their proposed act. In fact it appears that they appreciate the consequences better than anyone else does. The apotemnophiliac view of the matter is clearly that life will be better without the limb, whatever it takes. Carl Elliott spoke to some sufferers:

> "there is a simple, relentless logic to these people's requests for amputation. "I am suffering", they tell me. "I have nowhere else to turn." They realize that life as an amputee will not be easy. They understand the problems they will have with mobility, with work, with their social lives; they realize they will have to make

> countless adjustments just to get through the day. They are willing to pay their own way. Their bodies belong to them, they tell me. The choice should be theirs. What is worse: to live without a leg or to live with an obsession that controls your life? For at least some of them, the choice is clear which is why they are talking about chain saws and shotguns and railroad tracks." (Elliott)

There is a strong case to be made that apotemnophiliacs know exactly what they are doing when they choose to have a healthy limb amputated. If this is the case – if they are competent to make the decision to be harmed – then paternalistic legal intervention is not justified to prevent them entering into a contract with a doctor who will provide this service.

There is something disturbing about the desire of a person to inflict harm on his own body, or have such harm inflicted upon him, especially when the desire for harm is obvious and the harm desired extreme. When the harm is beyond "transient and trifling" even the liberal thinker might be sorely tempted to agree that the person's self-determination in this case may be set aside. After all, even Mill allowed that we may intervene for the protection of incompetents. In effect, (if it is to be interpreted as a case of justified paternalism) the law disallowing consent to bodily harm stipulates a threshold for competence based on the underlying values of self-determination and the good of the individuals involved. It says that the decision to consent to such a harm may be justifiably overlooked; the good of these people is too important for us to judge them competent in this regard. But the fact that we can call into question the incompetence of people who wish to have healthy limbs amputated shows that there is something fishy about this method of stipulating a threshold for incompetence.

If, as it seems, amputation is indeed the best possible outcome for apotemnophiliacs, it is a vivid example of what is wrong with the risk-related standards of competence proposed by Buchanan and Brock. Since the harm desired in this case is so extreme – the amputation of a limb – on their model apotemnophiliacs are likely to be judged incompetent to make the decision. It would be hard to deny that harm is being done, on our usual standards of what is harmful, even though the patients are satisfied with the post-operative state of affairs. But on all other grounds, including pragmatic, it seems that they should be given the responsibility for their own decision, provided they have sufficient mental capacity to make it. Competence is not a function of the riskiness of choices, and nor is the freedom that we are taking to hinge on the responsibility individuals take for their choices.

## 3.9 Conclusion

In medical practice and in medicine for the law, decision-making competence is what sets off those choices for which we are held responsible from those for which we need not or cannot take responsibility. As such, determinations of competence promise to provide the sort of separation between voluntary and non-voluntary choices that proved so elusive. Some models of competence are based on the riskiness of the choices involved, especially in borderline cases of competence, where thresholds must be set rather than discovered. But these models – and the model advocated by Buchanan and Brock in *Deciding for Others* in particular – seem to let hard paternalism in the back door. The decision as to whether an individual is competent to make a certain decision should not have anything to do with the riskiness of the decision, unless that riskiness makes it harder to comprehend the outcome of the choice. And it is equally absurd to suppose that riskiness has anything to do with the voluntariness with which a choice is made. These points were illustrated by the case of apotemnophilia, a proposed contract to have extreme harm performed upon one.

In current British law a person cannot consent to harm beyond the level of 'bodily harm', which is defined as harm more serious than merely transient and trifling. Apotemnophiliacs under this law may be prohibited from contracting with a doctor to have a limb removed. The basis for the law is paternalistic; it conforms to the definition outlined in Chapter 1. I argued briefly that if it is considered as an instance of paternalistic interference in a competent decision then it is unjustifiable – the law should not be concerned with prohibiting competent adults from carrying out their own life plans, "for their own good". To interpret the law as presumptively justified paternalism, we must read it as having stipulated a threshold of competence, which apotemnophiliacs (and all who wish to be harmed beyond a certain level) fail to meet: the desire for this level of harm is taken as constitutive of or overwhelming evidence for incompetence. In this stipulation the law sets a higher value on the good rather than the self-determination of those who wish to harm themselves.

But the incompetence of apotemnophiliacs can be called into question. It seems that in fact they *do* know better than anyone else what is good for them. They have a problem and there is currently only one solution: amputation is thus the best possible outcome at present for apotemnophiliacs. In stipulating a high threshold of competence based on valuing the good of apotemnophiliacs over self-determination, the law has failed to satisfy either. The

competence of apotemnophiliacs really has nothing to do with the extreme harm they wish to have done to them; they do not fail to be responsible for their actions simply because they are dangerous and have far-reaching consequences.

## Chapter 4

### 4.0 Introduction

Buchanan and Brock attempt to say exactly when we may justifiably decide for others. They claim that self-determination and the good of the individual concerned are values to be balanced, and that the exact nature of the balancing depends on features of the particular situation. But this kind of particularism about the ethics of paternalism seems likely to let hard paternalism in through the back door. That is, unless there is an independent theory of how self-determination and overall good should balance, it is open to those in positions of power to decide against the self-determination of individuals whenever they disapprove of, or fail to understand, the choice being made.

In this chapter I explore some normative theoretical questions which underpin the high evaluation of self-determination or sovereignty that is necessary for anti-paternalism at the level of legal theory. J.S. Mill vehemently rejects any imposition of the state on (what we now call) self-determination, when the choices in question are purely self-regarding, and Mill's *On Liberty* is a classic, complex and sustained argument against such impositions. But a puzzling aspect of Mill's philosophy is the apparent conflict between his principle of utility and the strong anti-paternalism he expresses in *On Liberty*.

Mill's consequentialism takes happiness to be the value to be maximised; the principle of utility says "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure" (Mill U 55). Mill's anti-paternalist stance is demonstrated when he says that the only justification for state intervention in the private life of a citizen is to prevent harm to others – "his own good, either physical or moral, is not a sufficient warrant" (OL 68). But *prima facie* there will be occasions where the increase in utility brought about by preventing someone from harming herself in fact outweighs that of allowing her to do as she pleases.

One response to this puzzle is to attempt to ground an extremely strong, if not absolute, ban on paternalism in a non-hedonistic consequentialism. In other words, happiness is not the sole

aim in the consequentialist weighing of right actions. Such a view might take self-determination to be one value among a number to be promoted. This position has intuitive merits, but may more resemble an intuitionistic pluralism like that of Buchanan and Brock than the robust consequentialism for which Mill aimed.

In the final part of this chapter I look at a different sort of consequentialism, that is intended to be applicable across all areas of government. Philip Pettit's republicanism is a consequentialist theory that takes dominion, a version of freedom, as its goal to be maximised. I explore the ramifications of this theory on legal paternalism. Though there are some problems remaining with the theory as a whole, Pettit's republicanism turns out to entail an absolute ban on paternalistic law.

In this chapter I will be focussing on possible consequentialist groundings for strong views against paternalism. This is partly because of the prevalence of consequentialist thinking in actual policy-making, and because consequentialist thinking seems legitimate for issues of law and politics, since these concern the welfare of a large number of people. Deontological reasoning, on the other hand, seems ideally suited to telling individuals how to treat others. The state seems to have duties beyond that of an individual, and may legitimately take liberties that individuals may not and possibly override principles that individuals may not. Thus, though it may be easy to frame this debate in terms of principles of autonomy and benevolence, it is necessary to understand how it can be framed in consequentialist terms, in order that the state as well as individuals have a strong theoretical underpinning to their actions.

## 4.1 Mill on paternalism

It's unlikely that Mill had anything like apotemnophilia in mind when he wrote the second part of his famous principle – that "the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. *His own good, either physical or moral, is not a sufficient warrant*" (OL 68) [my emphasis]. He probably was not considering extreme sadomasochistic violence, either. But Mill's arguments apply equally well to these practises as to whatever outrageous ways of living he had in mind when he wrote *On Liberty*. The sexual and physical extremes of recent times (and here I'm thinking, as well as of apotemnophilia, of bodily modification such as

having horns surgically implanted in one's head for the purposes of performance art) are simply variations on the age-old themes of subculture and eccentricity. They can be accepted as such because they harm no-one but those who engage in them, and (I would argue) benefit those who engage in them either enough to make up for harming them, or *through* harming them.

It is natural to conceive of the puzzle about paternalism in terms of the competing values of self-determination and beneficence, as Buchanan and Brock did. Beneficence demands that we prevent people from being harmed, yet a respect for their self-determination pulls us in the opposite direction. In the case of fully competent adults there should be a strong presumption in favour of self-determination – that is, a strong resistance to paternalism, which becomes sharpest when one considers interference in one's own choices, for one's own good. As Feinberg puts it, "the reader would not welcome having his own judgment overruled, or the "better values" of others substituted for his own" (Feinberg 62).

Mill's objection to paternalism is based on a number of strands of thought. We can first identify the one that is fairly straightforwardly utilitarian in character. First, Mill argues that a person is more likely to know his or her own good better than anyone else:

> "neither one person, nor any number of persons, is warranted in saying to another human creature of ripe years that he shall not do with his life for his own benefit what he chooses to do with it. He is the person most interested in his own well-being: the interest which any other person, except in cases of strong personal attachment, can have in it is trifling compared with that which he himself has; the interest which society has in him individually (except as to his conduct to others) is fractional and altogether indirect, while with respect to his own feelings and circumstances the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else" (OL 141-142).

Paternalistic interference, therefore, is likely to decrease the well-being of its object, because that person is more likely to know what is best for him.

Of course, since Mill's utilitarianism takes happiness (pleasure and the absence of pain) as its goal to be maximised, this argument is not straightforwardly utilitarian. To make it so, we must read 'well-being' as 'happiness' – so the individual is taken to know what will make him *happy* more than anyone else will. This can be granted without too much dispute. Paternalistic

interference will tend to decrease the happiness of those whose choices are infringed upon, because they are likely to know what makes them happy better than anyone else does.

The arguments from 'individuality' are a little more complicated. The exercise of choice, according to Mill, allows a person to develop into a more valuable human being. This person will display the qualities of observation, reasoning and judgment, activity, discrimination, and firmness and self-control, rather than those of "ape-like" imitation. "It is possible that he might be guided in some good path, and kept out of harm's way, without any of [the former qualities]. But what will be his comparative worth as a human being?" (OL 123) The person who has developed these personal characteristics by living out his or her own choices is Mill's ideal 'individual'.

The relationship between individuality and utility is complex. We can probably safely take Mill to hold at least that the utility of a society comprised of individuals has the potential to be greater than one comprised of 'sheep'. "As it is useful that while mankind are imperfect there should be different opinions, so it is that there should be different experiments of living." Further, Mill argues that "these developed human beings are of some use to the undeveloped - ...to those who do not desire liberty, and would not avail themselves of it" (OL128-129). Those few who are original are "the salt of the earth; without them, human life would become a stagnant pool." (OL 129) John Skorupski writes that "Much of what is most passionately felt in Mill's political philosophy is threaded on this strand – the idea of a society of human beings fully and variously developed, morally vigorous, self-determining." (Skorupski 338) All well and good – but in the context of utilitarianism as the doctrine of happiness-maximisation, this means that a society of individuals, or a mixture of individuals and others, will have a greater overall share of happiness than a more restrictive one.

Already we can see the difficulties Mill might have in reconciling the views expressed in *On Liberty* with those of *Utilitarianism*. The very potent rhetoric of *On Liberty* ("his *comparative worth* as a human being", for instance – what does *that* mean for a utilitarian?) defies the kind of analysis common in contemporary philosophy, rather lending itself to interpretations like Skorupski's that burden the meaning with even greater weight. Mill is propagandising in *On Liberty*, and it seems almost churlish to take him too literally. But there are important ideas in the text not explored elsewhere, and it remains to be seen whether they are consistent with those expressed in Mill's other works. Here we can flag the link between a society of

individuals and a happy society. It is a pleasant and optimistic thought that the two are closely linked, but one that might be rejected in favour of the idea that the society of individuals – the culturally and intellectually vibrant society – may be less happy. Further, its happiness might not be what we (or Mill, in the end) find valuable about it.

On the level of the individual, matters are still harder to reconcile with utilitarianism. The title of Chapter 3 – "Of Individuality, as one of the Elements of Well-Being" – immediately suggests that Mill took individuality as more than instrumental to a person's happiness. The following passage supports this view: "Where not the person's own character but the traditions or customs of other people are the rule of conduct, there is wanting one of the *principal ingredients of human happiness*, and quite the chief ingredient of individual and social progress" (OL 120) [my emphasis]. On a natural reading of this passage, Mill means to tell us that individuality is a component of happiness, and can therefore be included directly in the utilitarian calculus. But this quite clearly conflicts with Mill's hedonistic conception of happiness, stated elsewhere: "By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure" (U 55).

Mill's arguments that 'individuals' are necessary to the progression of society (with the subtext that the progression of society contributes to everyone's happiness) have some force. But the utilitarian value to the *individual* of individuality is less than clear. Firstly, in many cases individuals fail to know their own good better than someone else would. Health-care professionals will tend to know what is best for your health better than you yourself do; an older person may know what is best for your career (see Archard 1994). Your family may know what's best for your happiness – but even if you know what's best for your happiness, you might act otherwise than to promote it. It is quite easy to imagine a well-developed individual, in Mill's sense, who does not achieve much hedonistic pleasure. This example also works against a second possible reading of Mill's third argument – it is a counter-example to the view that happiness is the necessary result of becoming a fully-fledged individual. There is but a tenuous link between living autonomously and living a life of hedonistic pleasure.

It seems, therefore, that Mill's strong opposition to paternalism cannot be reconciled to his hedonistic consequentialism. The ethical goal of seeking to increase happiness overall is compatible with paternalistic interference in people's lives. People need not know best what conduces to their own happiness. Nor need self-development (for which liberty is a necessary

condition) produce happy individuals, though they may increase the happiness of others. With the connection between self-determination and happiness so tenuous, there will clearly be instances where to undermine the former is to increase the latter. For instance, it seems we will be justified in forcing therapy on a person who is moderately depressed, since this will make her (and potentially those close to her) happier. Moreover, hedonistic consequentialism justifies a *policy* of such intervention, since it would appear to promote happiness in every case to which it is applied. The recurring problem in reconciling the ideas expressed in *On Liberty* with *Utilitarianism* seems to be the narrow conception of the utilitarian goal advocated by the latter book. A strong regard for self-determination cannot be based on pleasure and the absence of pain as the ultimate goal of all humanity.

## 4.2 Beyond Happiness

A more substantive link, it has been argued, holds between self-determination and some more complicated notion of *well-being*. John Skorupski writes that Mill "would have been in a much stronger position if he had seen that the deep-seated response [against interference if no harm is being done to others] requires one to acknowledge autonomy as a human end which is not simply a part of happiness, but in its own right a distinct ingredient of well-being" (Skorupski 345). Skorupski goes on to argue that we need not give up a consequentialist basis for anti-paternalism – we simply need a richer, non-hedonistic consequentialism. Mill may have been heading in this direction with his analysis of the parts of happiness in *Utilitarianism.*

On this sort of view, it is not just happiness which is to be maximised, but some *complex* of human ends. If we take autonomy as part of that complex, we have a *prima facie* consequentialist case against paternalism. But we still do not have a strong anti-paternalism, for " autonomy is only one element of well-being, [so] the possibility remains open that an individual's overall utility may be increased by diminishing it. The cost of his loss of autonomy may be outweighed by the gain to other elements of his good" (Skorupski 359). Skorupski accepts this result - "Can one seriously deny that such cases do indeed exist?" (360) – but argues that paternalism will not be acceptable in the public sphere, that the "menacing indirect consequences of giving the state or society powers to interfere may indeed justify a ban on paternalistic moral practices or laws which is *in practise* absolute" (360).

The problem is that this variety of consequentialism does not even seem to ground the strong anti-paternalism which amounts to a practical ban. For autonomy is just one component, presumably among many, of well-being. Health must be another element of well-being. And health and autonomy are at odds whenever a competent person wishes to engage in some activity detrimental to her health. Competent persons choose to eat too much, drink too much, smoke cigarettes, work too hard, exercise too much, undergo plastic surgery, and consent to violent sexual activities. For some of these excesses it is clear that some other benefit than health is being promoted – it is easy to see how a person's well-being might be centred on his work, physical abilities, or sexual identity. But a competent person may even choose not to further *any* aspect of her well-being. This free spirit actively chooses a life where she fails to take account of the consequences of her action. On this view we will in many cases be justified in intervening in order to promote some other aspect of her well-being. The true altruist acts to promote some other person's good – on this view we are justified in preventing this if we can more greatly increase the altruist's well-being than he would have increased his recipient's. On a smaller scale, most of us are at least part altruist and part free spirit. These moments in which we choose to act against our own best interests are not aberrations to be prevented by some outside agency, but choices that contribute to our character just as do all our others. But a consequentialism which takes autonomy to be but one of a number of values to be maximised cannot sustain a case against such instances of paternalistic interference.

A natural response might be to try to give autonomy a higher place in the list of values to be maximised than any or many of the others. Thus, perhaps, we might be able to say that well-being is to be maximised, but only insofar as autonomy is not interfered with. However, this amounts to a second, non-consequentialist principle to add to that of maximising well-being – the view is no longer a pure consequentialism. This highlights a general problem with versions of consequentialism that take more than one value as their goal – like intuitionist positions in ethics, they can provide no satisfactory way of dealing with conflicts between goals. The catch-all term 'well-being' may promise a single goal for the consequentialist, but the reality is that well-being is a complex, whose elements *often* conflict. If this complicated consequentialism is not intuitionism in sheep's clothing, it at least faces the same deep problem of being unable to provide any system of ranking values.

## 4.3 Paternalism and freedom

In their 1976 paper, Bernard Gert and Charles Culver make the point that paternalism cannot be characterised in terms of a limitation on freedom. Their point is that there are acts which intuitively should fall under the concept of paternalism, and yet which do not limit freedom. They give the following example:

> "Mr. *N*, a member of a religious sect that does not believe in blood transfusions, is involved in a serious automobile accident and loses a large amount of blood. On arriving at the hospital, he is still conscious and informs the doctor of his views on blood transfusion. Immediately thereafter he faints from loss of blood. The doctor believes that if Mr. *N* is not given a transfusion he will die. Thereupon while Mr. *N* is still unconscious, the doctor arranges for and carries out the blood transfusion" (Gert and Culver 46).

Since the patient was unconscious, Gert and Culver argue, the doctor's act was not an imposition on the patient's freedom. This point was taken into consideration when framing our definition of paternalism, understanding it as Archard does in terms of usurpation of choice rather than limitation of freedom.

## 4.4 Two concepts of liberty

When Gert and Culver argue that limitation of freedom is not the defining feature of paternalism, they are dealing with a standard conception of freedom – "liberty of action" – which can be identified as Isaiah Berlin's 'negative' liberty. In an influential essay, Berlin argues that two conceptions of liberty have become entangled, and that only one is a fit concern of the state.

Berlin characterises negative freedom as follows:

> "I am normally said to be free to the degree to which no man or body of men interferes with my activity. Political liberty in this sense is simply the area within which a man can act unobstructed by others, If I am prevented by others from doing what I could otherwise do, I am to that degree unfree" (Berlin 393).

Freedom in this negative sense is often thought to be curtailed by law to some extent in order to promote some other values, for instance, justice, security, or equality.

> "This is what the classical English political philosophers meant when they used this word ['freedom']. They disagreed about how wide the area [of non-interference] could or should be. They supposed that it could not, as things were, be unlimited, because if it were, it would entail a state in which all men could boundlessly interfere with all other men; and this kind of 'natural' freedom would lead to social chaos in which men's minimum needs would not be satisfied; or else the liberties of the weak would be suppressed by the strong" (Berlin 393).

Where this conception of freedom is taken as paramount, the role of the state is to perform a balancing act. On the one hand, any imposition by law is curtailment of negative liberty, and to that extent is immediately in need of justification. On the other hand, without law there can be no harmony in the interests of individuals. Hobbes took this harmonisation of interests to justify a strong imposition on 'natural' liberty: only an absolute sovereign could save humanity from the constant state of war that is the result of their nature.

> "The final Cause, End, or Designe of men (who naturally love Liberty, and Dominion over others,) in the introduction of that restraint upon themselves, (in which wee see them live in Commonwealths,) is the foresight of their own preservation, and of a more contented life thereby; that is to say, of getting themselves out from that miserable condition of Warre, which is necessarily consequent (as hath been shewn) to the naturall Passions of men, when there is no visible power to keep them in awe, and tye them by fear of punishment to the performance of their Covenants" (Hobbes 117).

In short, "the greatest [worst] that in any forme of Governement can possibly happen to the people in generall, is scarce sensible, in respect of the miseries, and horrible calamities, that accompany a Civill Warre" (Hobbes 128).

Mill's rather more optimistic view of human nature and human endeavour led to a more minimalistic conception of the state. As we have seen, Mill put great emphasis on the value of liberty as instrumental to happiness, and happiness is the consequentialist goal of Mill's ethical utilitarianism. The ethical state, then, will take the maximisation of happiness as its goal, and concomitantly, liberty will be restricted to the smallest extent compatible with maximising happiness. Thus Mill's theory allowed the interference of the state by law *only* to prevent harm being inflicted by one citizen on another.

However, according to Berlin, negative freedom is not the only conception of freedom in the history of political thought. 'Positive' freedom is minimally construed as requiring the *presence* rather than the absence of some factor, though it is sometimes difficult to get a grasp on exactly what this additional factor might be. Berlin elucidates to some extent:

> "The positive sense of the word 'liberty' derives from the wish on the part of the individual to be his own master. I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other men's, acts of will" (Berlin 397).

However, it might be argued that the desire for negative liberty – freedom from interference – results from exactly the same wish in the part of the individual. The positive aspect of this sense of 'liberty' must be brought into relief. Berlin writes:

> "One way of making this clear is in terms of the independent momentum which the, initially perhaps quite harmless, metaphor of self-mastery... [The] dominant self is then variously identified with reason, with my 'higher nature', with the self which calculates and aims at what will satisfy it in the long run, with my 'real' or 'ideal' or 'autonomous' self, or with my self 'at its best'" (397).

This metaphorical language, Berlin argues, draws us into a state of mind whereby we can see a justification for all kinds of oppression on behalf of 'real' selves, on behalf, in fact, of the positive freedom of those we oppress. With characteristic vigour, Berlin writes that

> "This monstrous impersonation, which consists in equating what X would choose if he were something he is not, or at least not yet, with what X actually seeks and chooses, is at the heart of all political theories of self-realization. It is one thing to say that I may be coerced for my own good which I am too blind to see: this may, on occasion, be for my benefit; indeed it may enlarge the scope of my liberty. It is another to say that if it is my good, then I am not being coerced, for I have willed it, whether I know this or not, and am free (or 'truly' free) even while my poor earthly body and foolish mind bitterly reject it, and struggle against those who seek however benevolently to impose it, with the greatest desperation." [398]

This view takes paternalistic interference as non-coercive *because* it is for the person's own good. We saw (and rejected) earlier an attempt to justify paternalistic action in similar grounds – not only was paternalism taken to be for the recipient's good, it was taken to be ultimately their choice. The basis for this view was the same body of metaphor, of 'real' and 'base' selves, taken to the same literal extremes.

## 4.5 A third concept of liberty

The notion of liberty that Gert and Culver argue is not necessarily infringed by paternalistic intervention – the notion of liberty of action – is clearly identical with Berlin's concept of negative liberty. And it is correct to say that, in this sense, paternalistic interference need not involve an imposition on liberty. But it has been argued that Berlin's characterisation does not exhaust the possibilities of what we mean when we talk about liberty. In particular, Philip Pettit has argued that there is a third form of liberty that is even more worth pursuing than Berlin's negative liberty.

Pettit conceives of freedom as *non-domination*, the absence of mastery by another.

> "This conception is negative to the extent that it requires the absence of domination by others, not necessarily the presence of self-mastery, whatever that is thought to involve. The conception is positive to the extent that, at least in one respect, it needs something more than the absence of interference; it requires security against interference, in particular against interference on an arbitrary basis" (Pettit 1997 51).

According to Pettit, there are three components of any relationship of domination. The dominator has 1) the capacity to interfere 2) on an arbitrary basis 3) in certain choices that the other is in a position to make (52). The absence of domination so characterised is freedom as non-domination.

Interference is characterised as being intentional, and (for Pettit at least) negative: "Interference cannot take the form of a bribe or reward; when I interfere I make things worse for you, not better" (52). I am sceptical about this characterisation of interference, but let's leave it for the moment – the problem will come out when we talk about arbitrariness. Pettit's focus on the *capacity* of the dominator to interfere brings out one clear difference with the notion of negative liberty. For a person to be unfree in the sense of negative liberty, they must be interfered with. But for a person to be unfree in the sense of non-domination, there need only be the capacity for someone to arbitrarily interfere in their choices – this is Pettit's 'non-interfering dominator'.

> "What constitutes domination is the fact that in some respect the power-bearer has the capacity to interfere arbitrarily, even if they are never going to do so. This fact means that the power-victim acts in the relevant area by the leave, explicit or implicit, of the power-bearer; it means that they live at the mercy of that person, that

> they are in the position of a dependent or debtor or something of the kind. If there is common knowledge of that implication, as there usually will be, it follows that the power-victim cannot enjoy the psychological status of an equal: they are in a position where fear and deference will be the normal order of the day, not the frankness that goes with intersubjective equality" (63-64).

For Pettit, the threat of interference can itself constitute unfreedom. For freedom to be realised, interference must be impossible, not merely unlikely.

But there is another striking difference between freedom as non-domination and negative freedom, brought out by the second clause in Pettit's definition of domination. That is that just as there may be a non-interfering dominator who reduces freedom as non-domination, so may there be a non-dominating interferer, who does *not* reduce freedom as non-domination. For the negative libertarian, every interference is an imposition on freedom. But for Pettit, only *arbitrary* interference (or the capacity to arbitrarily interfere) imposes on freedom as non-domination. Ideally, the state will be an example of a non-dominating interferer, having the capacity to interfere in the lives of individuals but not on an arbitrary basis. It is therefore crucially important to understand what Pettit means by non-arbitrary interference.

> "When we say that an act of interference is perpetrated on an arbitrary basis…we imply that like any arbitrary act it is chosen or not chosen at the agent's pleasure. And in particular, since interference with others is involved, we imply that it is chosen or rejected without reference to the interests, or the opinions, of those affected. The choice is not forced to track what the interests of those others require according to their own judgements" (55).

A non-arbitrary interference will track the interests – on their own estimation – of the person interfered with, "Or…[is] at least forced to track the relevant ones" (55).

There is a certain degree of tension between the ideas of freedom as non-domination and an extensive welfare state, both of which Pettit wishes to support. He needs to ensure that the interference required by an extensive state is non-dominating. To be non-dominating, the interference must be non-arbitrary, and at times the notion of non-arbitrariness seems to be stretched to accommodate this need.

> "I may have an interest in the state imposing certain taxes or in punishing certain offenders, for example, and the state may pursue these ends according to procedures that conform to my ideas about appropriate means. But I might still not want the state

> to impose taxes on me – I may want to be an exception – or I may think that I ought not to be punished in the appropriate manner, even though I have been convicted of an offence. In such a case, my relevant interests and ideas will be those that are shared in common with others, not those that treat me as exceptional, since the state is meant to serve others as well as me. And so in these cases the interference of the state in taxing or punishing me will not be conducted on an arbitrary basis and will not represent domination" (55-56).

But what of the person who genuinely (though erroneously) estimates their own interests as being free of taxation? Is this person dominated under Pettit's system of government? Not, according to Pettit, if there are effective channels for them to contest the law.

> "Happily, a little reflection shows that what is required for non-arbitrariness in the exercise of a certain power is not actual consent to that sort of power but the permanent possibility of effectively contesting it. [...] Unless such contestability is assured, the state may easily represent a dominating presence for those of a certain marginalized ethnicity or culture or gender" (63).

I am not entirely convinced that a person with interests different from those of the majority will still not be dominated under Pettit's system, but the important thing for us to note is that the focus on a person's interests *in their own estimation* has been maintained: Pettit's state, though it allows interference in the interests of its citizens, is not perfectionistic.

## 4.6 Freedom and consequentialism

Pettit, like Mill, is a philosophical consequentialist. But where Mill takes maximising happiness to be the goal of the ethical state, Pettit believes that the state should maximise freedom as non-domination. We saw earlier that there was some difficulty with a consequentialist theory sustaining a robust anti-paternalism – one might think that a theory of constraints rather than a theory of goals is more suited to the task. Pettit's consequentialism, with its goal of non-domination, at least promises to do better, for domination seems to be related to paternalism in way that simple interference is not (on this more later.)

Pettit, with John Braithwaite, defends a consequentialist approach (in this case, for the criminal justice system) in *Not Just Deserts*. Braithwaite and Pettit argue that *constraints* on an agent's behaviour (such as 'do not punish the innocent'), usually associated with

deontological theories, can be derived from their consequentialist theory of punishment, which takes 'dominion' (an earlier version of non-domination) as its goal.

> "It ought to be clear that the target will motivate a legal allocation of uncontroversial rights. If dominion is to be promoted by legal sanction, then certain negative liberties must certainly be legally protected. Their protection means that citizens will have a legal claim on the state to defend such liberties. More specifically, given how the law works, it means that they will have a claim which legally constrains the state rather than just providing it with a target: the state will not be able to excuse inaction on the grounds that this will better serve the defence of that sort of right overall, or the like" (Braithwaite and Pettit 71).

It is the nature of the goal that allows this derivation of rights. Where the consequentialist goal is happiness, or welfare, a state representative might always ask if violating a derived right on this occasion will not promote the good overall. But (Braithwaite and Pettit argue) dominion is such that an individual violation of a right cannot be countenanced:

> "it is a goal whose promotion requires the agents of the system individually and collectively to tie their hands in regard to how individuals should be treated and to make it clear to people that this is what they are doing. Unless they make such a commitment then people generally will become aware that they are likely to have their personal dominion invaded if that is for the best overall; and since dominion has the subjective dimension, this means that people generally will find that their dominion is seriously compromised" (73-74).

Pettit develops this idea further in *Republicanism*. He argues that dominion has a *common knowledge* feature: when the three conditions of domination are fulfilled in any degree, any one party to the relationship is likely to know about it: "The conditions may not be articulated in full conceptual dress, but the possibilities involved will tend to register on the common consciousness" (59).

But *non-domination* as opposed to *dominion* as a goal cannot use this common knowledge feature to prove that constraints derived from it cannot consistently be violated. Though there is a subjective element to domination, it is in no way constitutive of domination (as Pettit conceives of it). Domination is characterised by the capacity to interfere on an arbitrary basis in the choices of the dominated. Any violation of the rights or constraints derived from the target of non-domination will reduce domination, but if it reduced domination by a greater degree in some other area, why not violate the right in this instance? For example, if one

person is dominating a number of other people, for instance in an employment situation, why not arbitrarily interfere to prevent this domination. The answer seems to be in the feature of domination that is the *capacity* to interfere. For the state that has the capacity to interfere in that employer's life on an arbitrary basis would seem to have the capacity to interfere in anyone's life. And so, by the definition, the state which can interfere on an arbitrary basis, *even in an attempt to promote non-domination*, actually dominated every citizen. Thus Pettit's view of non-domination as a consequentialist target supports constraints on the behaviour of agents to *protect* their freedom as well as promote it.

## 4.7 Paternalism and domination

When writing from the perspective of *dominion* as a consequentialist target, rather than *non-domination*, Braithwaite and Pettit note that legal paternalism may be justified. Regarding heroin use, they write:

> "In the long term, addiction will reduce the dominion of the consensual victim. Ultimately, it may give her no choice but to run every aspect of her life to service the habit; it may leave her with no resources to resist the manipulations of dealers who use her money, or pimps who use her body. This means that there is a case to consider in favour of criminalizing heroin use" (97).

Braithwaite and Pettit go on to argue that for reasons of parsimony, this sort of case is unlikely to be a matter for the criminal justice system (93-94; 97-98). But the fact that there is a case for the criminalisation of heroin use on the dominion model shows that dominion is not the version of negative liberty that Braithwaite and Pettit take it to be.

Braithwaite and Pettit explain that

> "Negative liberty involves, roughly, the absence of interference by others, or at least of interference that is sufficiently intentional, negligent, reckless, or indifferent to count as blameworthy or culpable...Positive liberty involves something more in addition. What more is required varies from one account to another. It may be the absence of physical inability, psychological incapacity, personal ignorance or something of that kind" (Braithwaite and Pettit 55).

They take dominion to be a form of negative liberty.

> "The key to understanding the rival republican tradition of interpretation is to see that there is an alternative way of characterising non-interference. The condition, under

> the republican interpretation, is this: that there are others around, and they do not interfere. The place of the negation operator is different in this characterization, for it is now required that there are others around who do not interfere; it is not enough if there are no others available to interfere" (57).

But if the case of drug addiction is one where dominion is lost, then dominion must have an element of positive liberty – the only way that a drug addict's dominion could be reduced is if an additional psychological element is required for full dominion (this is not true: they say it's about the dealers and pimps – but perhaps this is farfetched).

Freedom as non-domination seems to be a less ambiguous goal for the consequentialist to pursue. The drug addict is not dominated – at least not by definition. Her dealer may well be in a position of dominating power, but that is no grounds for criminalizing the *use* of addictive substances. What we want now to assess is whether the criminalisation of drug use – the paternalistic intervention of the state – is an instance of domination, and to what extent Pettit's model supports an anti-paternalistic stance.

Whether or not paternalism – the usurpation of one's choices *for one's own good* – counts as domination depends on the crucial arbitrariness clause. If the interference is arbitrary then paternalistic law goes against the consequentialist goal of freedom as non-domination. But paternalistic law *is* often arbitrary, provided we stick with Pettit's characterisation of non-arbitrariness as tracking the interests of those affected *on their own estimation.* For paternalism tends to involve usurping a choice made by an individual on their own estimation, in order to promote (what are felt to be) their actual interests. If arbitrariness involves failing to track a person's interests on their own estimation, then paternalism is often arbitrary intervention in a person's choices. As such it is domination.

But it seems that paternalism need not always be domination – the two are conceptually separable. Paternalistic intervention involves the usurpation of a person's choices. Often this will mean that their interests are ignored. But it is conceptually possible for paternalistic intervention to be non-arbitrary interference in a person's choices – an interference that *does* track their interests, in their own estimation. It seems that the smoker who wants to quit, who fervently assures her friend that she wants to quit, and persuades him to look after her money on a night out so that she can't buy cigarettes, is not *dominated* when she later begs for cash, though she is the subject of paternalistic intervention.

## 4.8 Paternalism and domination II

In Part I, we viewed paternalism as being (in act or in law) motivated by benevolence and in tension with self-determination. The idea of freedom as non-domination adds an extra element to this picture. It may give us a new way of understanding what is wrong with paternalism when it is wrong, and why it is sometimes not wrong. On our earlier model, paternalism was thought to be wrong insofar as it denied a person's self-determination. But paternalism denies self-determination by definition, since we argued that most choices have to be seen as expressions of self-determination. Therefore this model failed to give us adequate grounds for understanding what is wrong with paternalism when it is wrong, or why it is justified when it seems to be.

What if, instead, we understand paternalism as sometimes instantiating a relationship of domination – sometimes demonstrating the capacity of a person or state to interfere arbitrarily in another's choices? As we saw above, many instances of paternalism are clearly such cases. Whenever the interests of the subject – in her own estimation – are not consulted, paternalism fits the model of domination. It is arbitrary, arbitrariness being defined in terms of non-consultation of interests, and by the definition of paternalism it is interference in the choices of its object.

Understanding paternalism as being wrong when it instantiates a relationship of domination allows us to recognise that cases where individuals *ask* for their later choices to be usurped – as in the smoking example above – are unproblematic. Though the later choice of the smoker is interfered with, and to a certain extent her self-determination is interfered with, her interests in her own estimation have been consulted, and so the case is not one of domination. We can also see, perhaps, why such wildly differing intuitions are generated when we consider paternalism from two perspectives – that of the interferer and that of the interfered-with. While the person whose choices have been usurped without reference to his stated interests feels diminished, the usurper may feel aggrandised. She, after all, is put in a position of power with respect to the other person. Best of all, she can put it all down to benevolence, since her action is 'for his own good.'

But where does this leave us on the issue of competence? Many thinkers have agreed on the idea that paternalism is justified where people are incompetent to make their own decisions.

Are such instances of paternalism also instances of domination? Take first the easier case, where an individual is clearly incompetent to make a decision – a person suffering from serious senile dementia, for example. In this case, a person may not even have an estimation of her own interests. If she has a living will or other documentation stating her interests, this must be consulted lest interference in her life be arbitrary and therefore dominating. But if she has no stated interests then her carers must do the best they can – and it seems unfair to say that they dominate her simply because they fail to account for her non-existent statement of her own interests.

A more difficult case is one where a person who is clearly incompetent has firm interests that go against his well-being. This is the case where, traditionally, paternalistic interference has been thought to be justified. Take the young man, a schizophrenic, who has become suicidal. Should he be hospitalised until he seems stable again? If the interference in his choices is arbitrary, then a relationship of dominance is established, it seems. Though I am hesitant to add more complexity to the definition of non-arbitrariness, we might say that the stated interests of this individual may justifiably be ignored, simply because he is incompetent to make decisions. (Compare the case of a will made when in unsound mind.)

This brings us back to a point made in my third chapter. That is, that if determinations of incompetence are to have such a profound effect on our ethical judgements – if actions which are standardly thought to be unjustified can become justified when the person they are applied to is judged to be incompetent – then determinations of incompetence themselves must be made independently of the considerations we take into account when judging the rightness or wrongness of the action. A serious problem with the model explored in the previous section was that it made determinations of competence rest on the very factors that we were trying to use to decide whether paternalism was right or wrong. What made paternalism wrong was that it was a violation of self-determination. But where a person was borderline-incompetent, if their self-determination was simply judged to be unimportant in comparison with their well-being, then they were judged incompetent. This is a backwards sort of reasoning.

If we can judge a person as incompetent based on their cognitive abilities, this judgement can be used to justify ignoring their choices. We can say that interference in persons' choices, even if it goes against their stated interests, is not arbitrary *if they are incompetent to make the decision*. But our judgment of their incompetence cannot be based on our simply not liking

that decision, or thinking that it so goes against their well-being that we are justified in ignoring their self-determination. That sort of judgment of incompetence really does seem to be arbitrary – in the sense of being at the whim of the decision-maker.

## 4.9 Legal paternalism

On this model of unjustified paternalism as domination, most if not all instances of legal paternalism will be instances of domination. To prevent people from harming themselves by instituting laws against self-harming behaviour will in many if not most instances be to interfere arbitrarily in their choices. But on Pettit's model an absolute ban on legal paternalism can be derived. For Pettit, the goal of the state is to maximise non-domination. The state must therefore be the apotheosis of non-domination. But since domination is not only arbitrary interference but the *capacity* to arbitrarily interfere, there will be an absolute ban on paternalism: laws designed to prevent competent people from engaging in self-harming behaviour give the state the *capacity* to interfere in the choices of competent individuals, without reference to their interests in their own estimation.

## 4.10 Conclusion

We can't accept the general argument that self-determination is instrumentally crucial to a people's well-being because they know their own good more often than does anyone else; often people do *not* best know their own good. But we can accept Mill's cautions about the danger of using the law in any situation where self-determination is at stake. We cannot, either, side with Feinberg and argue that sovereignty is a trumping consideration. We saw, from Buchanan and Brock's argument, that other aspects of well-being refuse to be ignored. But we can argue that self-determination is partly constitutive of well-being, and the two should not be seen as opposed. Rather, different *elements* of well-being may be opposed.

However, from the perspective of underlying normative theory, these conclusions are problematic. Mill's anti-paternalistic stance is apparently in conflict with his hedonistic consequentialism, and expanding the target of the consequentialism to a complex notion of well-being doesn't help. In simple form this view can't ground a strong anti-paternalism, but in complex form it faces problems of ranking elements of well-being. A consequentialism that takes a complicated form of well-being as its goal faces the same difficulties that face

intuitionism. The reasons we gave for valuing self-determination over other elements of well-being would have to be introduced as additional principles to the single principle of maximising well-being.

Since Gert and Culver gave the example of the unconscious man's preferences being overridden by the paternalistic doctor, paternalism has been kept conceptually separate from freedom of action: a paternalistic act need not be one in which freedom is diminished. The notion of freedom in play here is Isaiah Berlin's negative liberty, characterised as the absence of interference by other people. Berlin contrasts negative liberty with positive liberty, understood rather vaguely as requiring the *presence* of some additional factor, rather than the mere absence of interference. Berlin describes this additional factor as 'self-mastery'.

But, according to Philip Pettit, Berlin's two concepts of liberty do not exhaust the field, and negative liberty does not exhaust the field of liberties worth pursuing. Pettit describes a third concept of liberty – liberty as non-domination – and sets it up as the consequentialist goal of a 'republican' government. The relationship of domination involves the capacity of one party to arbitrarily interfere in the choices of another, where arbitrary interference is understood as interference without consultation of the interests of the second party, in their own estimation.

Paternalism and domination are clearly closely related, both involving interference with the choices of agents. Paternalism need not be dominating, since it need not involve *arbitrary* interference. Nor, of course, need domination be paternalistic – domination need not be for a person's 'own good', for one thing, and domination may involve only the *capacity* to interfere, while paternalism involves actual interference. But I have argued that the perspective of domination gives us a better understanding of when paternalism is justified and when it is unjustified. Paternalism which instantiates a relationship of domination is unjustified.

This model explains why the positions of the intervener and the intervened-with are so at odds. It also allows us to say that some paternalistic interference with the choices of competent individuals is not wrong, provided it goes along with their expressed interests. It helps us to understand why the previously expressed interests of incompetent patients are so important – they guarantee that interference based on these interests is non-dominating. As for determinations of competence, it has become clear that they must be made independently of

judgments of the justifiability of paternalism. Once it is determined that a patient is incompetent, it seems we can fail to account for their expressed preferences without arbitrariness, though great care will be required in determinations of competence to prevent domination.

Finally, on Pettit's model of non-domination as the consequentialist goal of the republican state, legal paternalism can be absolutely banned. For a state which has laws preventing harm to self has the *capacity* to arbitrarily interfere in the lives of its citizens, and this counts as domination. The only paternalistic intervention the state is allowed, in fact, is to fail to uphold a contract whereby one party sells himself into slavery. This seems to be the only state action which, while (perhaps) dominating itself, prevents domination on a larger scale: "by selling himself for a slave, he abdicates his liberty; he foregoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself" (Mill *On Liberty* 173). Mill is generally taken to be speaking of a variety of negative liberty, but his words apply just as well, if not better (since a master need not *actually* interfere with the freedom of action of her slave), to liberty as non-domination.

## Conclusion

### 5.0

In this thesis I have taken the question of the justifiability of legal paternalism as a starting point for explorations of ideas of freedom and responsibility at a number of levels. I started by defining paternalism and setting out the two guiding intuitions of the thesis, that paternalistic intervention in the lives of competent adults is unwarranted, but may be justified in cases of mental incapacity, ignorance, or youth.

These intuitions were given a preliminary formalization in Joel Feinberg's theory of legal paternalism. According to Feinberg, paternalism is justified where an action is substantially non-voluntary, because non-voluntary actions are not expressions of the trumping right to self-determination. Paternalistic intervention in the lives of competent adults is unwarranted because of this right to sovereignty. But cases of mental incapacity, ignorance and (sometimes) youth are cases of non-voluntary action, and paternalism may be justified.

This exploration of Feinberg's theory demonstrated that there are two aspects to the guiding intuitions of the thesis: the non-ethical or metaphysical aspect, and the ethical aspect. The non-ethical aspect requires a division of actions or choices into two classes – the voluntary and the non-voluntary – while the ethical aspect imparts a value to one of the two classes that the other lacks. My second chapter explores the non-ethical aspect of the intuitions, while the third and fourth deal with the ethical.

A problem with Feinberg's theory of when paternalism is justified emerged in my second chapter. That is, that there is some difficulty in distinguishing voluntary from non-voluntary choices. Feinberg's response to this problem is consistent with more than one theory of freedom of the will. I rejected Harry Frankfurt's account as requiring an unrealistic conception of individual identity, opting instead to base further investigation on Dennett's idea of freedom as created by practises of attributing responsibility.

In my third chapter I started to look at some of these practises of attributing responsibility. Theories of decision-making competence such as that of Buchanan and Brock attempt to set out exactly when it is appropriate to interfere in the decisions of individuals. Competence determinations are therefore in a sense the creators of freedom and personal identity, since they are the practical element required by Dennett's theory. But competence determinations may fail on various points, and Buchanan and Brock's method fails on one extremely important point. That is, the determination of competence on their account rests on a prior evaluation of the relative weights of the self-determination and the good of the person whose competence is in question. But their capacity to make decisions is independent of these factors, and to use them as the basis of the competence determinations risks hard paternalism all over again.

How, then, should we act in the practical arena when it comes to paternalistic intervention? My fourth chapter explores possible normative ethical bases for a strong presumption against paternalism in the law. I start by looking at the classic anti-paternalist statement – Mill's *On Liberty* – and pointing out the difficulty of reconciling Mill's views on paternalism with his utilitarianism. In short, it seems impossible to obtain a high enough valuation of self-determination when happiness as 'pleasure and the absence of pain' is one's consequentialist goal. Skorupski interprets Mill's consequentialism as having some more complex goal, but this strategy risks some kind of pluralism (with which Mill would have been most displeased.)

Finally, we investigated how the law would be affected with respect to paternalism on a consequentialist model that takes freedom as non-domination as its goal. Valuing freedom as non-domination gives a normative basis for rejecting paternalism in all cases except where incompetence has been independently determined.

## References


Archard, D. (1990) 'Paternalism Defined', *Analysis* 50: 36-42

—— (1994) 'For Our Own Good', *Australasian Journal of Philosophy* 72: 283-293

Aristotle (2000) *Nicomachean Ethics* Cambridge: CUP

Berlin, I. (1997) 'Two Concepts of Liberty', in Goodin, R.E. and Pettit, P. *Contemporary Political Philosophy: an anthology* Oxford: Blackwell

Braithwaite, J. and Pettit, P. (1990) *Not Just Deserts: A Republican Theory of Criminal Justice* Oxford: Clarendon Press

Buchanan, A.E. and Brock, D.W. (1989) *Deciding for Others* Cambridge: CUP

Cabanne, P. (1975) *Van Gogh* Translated from the French by Daphne Woodward. London: Book Club Associates

Darwall, S. (1998) *Philosophical Ethics* Oxford: Westview Press

Dennett, D. C. (2003) *Freedom Evolves* London: Allen Lane

Dworkin, G. (1988) *The Theory and Practise of Autonomy* Cambridge: CUP

Elliott, C. (2000) 'A New Way to be Mad', *The Atlantic Monthly Review* http://www.theatlantic.com/issues/2000/12/elliott.htm.

Feinberg, J. (1970) 'Causing Voluntary Actions', in *Doing and Deserving: Essays in the Theory of Responsibility* Princeton: Princeton University Press 152-186.

—— (1986) *Harm to Self* New York: OUP

Frankfurt, H. (1971) 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68: 5-20.

—— (1988) *The Importance of What we Care About* Cambridge: CUP

Gert, B. and Culver, C. (1976) 'Paternalistic Behaviour', *Philosophy and Public Affairs* 6: 45-57.

Haji, I. (2002) 'Compatabilist Views of Freedom and Responsibility', in Kane, R. (ed) *The Oxford Handbook of Free Will* Oxford: OUP 202-228.

Hart, H.L.A. (1963) *Law, Liberty and Morality* Oxford: OUP

Hobbes, T. (1996) *Leviathan* Cambridge: CUP

Kell, D. (1994) 'Social Disutility and the Law of Consent', *Oxford Journal of Legal Studies* 14 (121-135)

Mill, J.S. (1912) *On Liberty Etc.* London: OUP

—— (1998) *Utilitarianism* Oxford: OUP

O'Neill, O. (2002) *Autonomy and Trust in Bioethics* Cambridge: CUP

Pettit, P. (1997) *Republicanism* Oxford: Clarendon Press

Riley, J. (1998) *Mill on Liberty* London: Routledge

Van Inwagen, P. (1975) 'The Incompatibility of Free Will and Determinism', *Philosophical Studies* 27 (185-199).